Computers & Industrial Engineering 137 (2019) 106033

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

Simulation of an automotive supply chain using big data

António A.C. Vieira[a,*], Luís M.S. Dias[a], Maribel Y. Santos[b], Guilherme A.B. Pereira[a], José A. Oliveira[a]

[a] Department of Production and Systems, ALGORITMI Research Centre, University of Minho, Portugal
[b] Department of Information Systems, ALGORITMI Research Centre, University of Minho, Portugal

ARTICLE INFO

Keywords:
Simulation
Big Data
Supply chain
Risks
Industry 4.0

ABSTRACT

Supply Chains (SCs) are dynamic and complex networks that are exposed to disruption, which have consequences hard to quantify. Thus, simulation may be used, as it allows the uncertainty and dynamic nature of systems to be considered. Furthermore, the several systems used in SCs generate data with increasingly high volumes and velocities, paving the way for the development of simulation models in Big Data contexts. Hence, contrarily to traditional simulation approaches, which use statistical distributions to model specific SC problems, this paper proposed a Decision-Support System, supported by a Big Data Warehouse (BDW) and a simulation model. The first stores and integrates data from multiple sources and the second reproduces movements of materials and information from such data, while it also allows risk scenarios to be analyzed. The obtained results show the model being used to reproduce the historical data stored in the BDW and to assess the impact of events triggered during runtime to disrupt suppliers in a geographical range. This paper also analyzes the volume of data that was managed, hoping to serve as a milestone for future SC simulation studies in Big Data contexts. Further conclusions and future work are also discussed.

1. Introduction

Supply Chains (SCs) are complex and dynamic networks, comprised by entities, such as suppliers and customers, wherein material and information exchanges occur, driven by demand and supply interactions (Shah, Chaudhari, & Cárdenas-Barrón, 2018; Simchi-Levi, Kaminsky, Simchi-Levi, & Shankar, 2008). Several activities take place in these networks, e.g., production, transport of raw materials from supplier to manufacturers. Ultimately, the goal of each entity in these networks may be seen as the fulfillment of their customers' orders, at a minimum cost, whilst improving their competitiveness. In other words, to efficiently manage raw materials receipt and timely schedule deliveries at the right time, place and quantities (Azzi, Chamoun, & Sokhn, 2019).

Supply Chain Management (SCM) plays a relevant role in managing SCs (Yin, Nishi, & Zhang, 2016). To ensure its efficiency, Levi et al. (2008) emphasized three different aspects of SCM. Firstly, the need for delivering products according to customers' requirements, thereby, in some SCs, it may be necessary to account for suppliers' suppliers and customers' customers. The second aspect emphasized by the authors is related with reducing total costs throughout the entire SC. This includes, but is not limited to, transportation and inventory costs (Gharaei, Karimi, & Hoseini Shekarabi, 2019a; Giri & Masanta, 2018). Finally, the authors also emphasized the importance of integrating all agents in the SC. These aspects are of utmost importance, since SCs evolve over time, wherein relationships between entities, such as customer demand, supplier capacity, among others, may abruptly change with possible negative consequences.

The highly stochastic nature, inherent to SCs, originates risks, caused by uncertain disruptions and variability, which may lead to a decrease in the involved entities' service level (Bottani, Murino, Schiavo, & Akkerman, 2019). To handle these types of problems, Supply Chain Risk Management (SCRM) emerged as a recent discipline (Er Kara, Oktay Fırat, & Ghadge, 2018). Usually, these disruptions originate from either the customer or the supplier end of the SC. For instance, customers' demand variability may originate situations in which production is insufficient, leading to orders not fulfilled, or situations where materials are stocked, because they are no longer required by customers. Furthermore, disruptions that are rare but have considerable negative impacts may also occur. These can be of short-term (e.g. transportation delay) or of long-term nature (e.g. earthquake that affects the production of a supplier) and may jeopardize crucial SC activities, such as production, which, in its turn, may also lead to

* Corresponding author.
E-mail addresses: antonio.vieira@dps.uminho.pt (A.A.C. Vieira), lsd@dps.uminho.pt (L.M.S. Dias), maribel@dsi.uminho.pt (M.Y. Santos), gui@dps.uminho.pt (G.A.B. Pereira), zan@dps.uminho.pt (J.A. Oliveira).

https://doi.org/10.1016/j.cie.2019.106033

Available online 31 August 2019
0360-8352/

customers' orders unfulfillment and hence negatively affect the SC. See the work of Ho, Zheng, Yildiz, and Talluri (2015) for a proposal of a classification of the types of risks that occur in a SC, which was developed after reviewing related literature.

Several real cases of disruptions in SCs can be found in literature (Giri & Bardhan, 2014). One example was provided by Matsuo (2015) and Park, Hong, and Roh (2013), who described the consequences of an earthquake that occurred in March 2011, which damaged the Tohoku Pacific Coastline, causing severe damage to factories in that zone, while also affecting other companies, including those not directly affected by the earthquake. According to Matsuo (2015), most consequences were related to the fact that all the first-tier suppliers purchased the same materials from the same factories, which were damaged by the earthquake. Because of this, when these non-first-tier suppliers failed to fulfill orders, several other plants were affected. This problem is prolonged in time, since usually suppliers take time to recover from these failures, due to constraints, such as their production capacity.

A proper assessment of the consequences of the above mentioned disruptions in a SC is not easy to determine. Thus, rather than being preemptive, companies often adopt reactive approaches, due to the lack of proper tools. In light of this, simulation approaches can be used as proactive Decision Support Systems (DSS) in SCs, for several purposes, including: testing alternative scenarios, prediction, understanding the behavior of complex systems, determine certain performance measures, or simply to animate the logistic flows, allowing to discover new knowledge from raw data. However, nowadays, SC systems generate data at increasingly higher rates, volumes and formats, in what is known as the three main characteristics of Big Data (Zikopoulos & Eaton, 2011), In fact, according to Madden (Madden, 2012), this environment in which data is too big, too fast and too hard for existing tools to process, paved the way for the advent of alternative structures to store and integrate data in Big Data contexts. As such, several studies have been emphasizing the need to couple simulation with Big Data concepts and technologies (Kagermann, Helbig, Hellinger, & Wahlster, 2013; Lasi, Fettke, Kemper, Feld, & Hoffmann, 2014; Tiwari, Wee, & Daryanto, 2018; A. A. Vieira, Dias, Santos, Pereira, & Oliveira, 2018; Zhong, Newman, Huang, & Lan, 2016).

In light of this, such tool is currently being developed at a Bosch Car Multimedia plant, which produces automotive electronics components. This tool consists of a Big Data Warehouse that supports the SC simulation model by storing, integrating and providing real data to it. The research that was conducted to develop part of the BDW has already been published (Vieira, Pedro, Santos, Fernandes, & Dias, 2018). Furthermore, a prototype of the simulation model was already published in (A A C Vieira et al., 2018). This latter did not use real industrial data, but rather used the typically available approach in simulation, consisting in using random distributions, to validate the data model, i.e., the variables selected for the project.

In light of the above, the purpose of this paper is threefold. First, it proposes a SC simulation model that receives data from the BDW and reproduces the associated material and information flows, working as a virtualization of the respective SC, as per the Industry 4.0 agenda (Kagermann et al., 2013; Lasi et al., 2014). Second, the paper discusses some aspects related with the volume of data handled for this research, thus, serving as a milestone for future SC simulation studies in Big Data contexts. Third, it shows how the tool can be used to enhance the decision-making process in Big Data contexts. For this purpose, the tool is used to test risk scenarios that mimic the behavior of the real system (using the data stored in the BDW) and, at the same time, are able to assess the impact of geographic disruptions.

This paper is structured as follows. Section 2 analyzes literature related with (1) simulation studies that used structures to provide data to SC simulation models and (2) existing structures used to store and integrate data in Big Data environments. Section 3 describes the methodology applied in this research, starting with a description of the SC system and followed by the approach that was defined, which includes the description of the proposed framework. Section 4 starts by providing some details regarding the development of the simulation model. The section also analyzes the volume of data managed by the implemented first instance of the BDW. Section 5 displays and discusses the results that can be achieved with the proposed simulation tool. In its turn, Section 6 analyzes the managerial implications of this research. Finally, Section 7 discusses the main conclusions, identifies the main scientific contributions and limitations of this paper, and also identifies future research directions.

2. Related work

The need to improve industrial processes is, in fact, one of the main goals of Industry 4.0 as is emphasized by Kagermann et al. (2013). Such improvement may involve several methods, with literature concerned with quantitative models in SCs being vast. For instance, Hoseini et al. (2018) proposed an optimization model that considered limited warehouse space to the optimization problem. Sarkar and Giri (2018) determined the optimum ordering parameters that minimize the total joint costs between vendor and buyer. Gharaei, Hoseini Shekarabi, and Karimi (2019) aimed to minimize the total inventory costs while maximizing the profit at the same time. Gharaei, Karimi, and Hoseini Shekarabi (2019b) proposed an optimization model to determine the optimal joint-economic lot size and agreed period lengths to minimize the total inventory costs in the SC, with stochastic constraints. Giri and Masanta (2018) considered the learning factors of human operators in a closed-loop SC.

Notwithstanding the important contributions of the cited, as well as other relevant studies, most do not model the entire SC and do not consider the detail that using real industrial data provides, as corroborated by (Jahangirian, Eldabi, Naseer, Stergioulas, & Young, 2010; Pires et al., 2016). Thus, it has been widely accepted that there is a need for studies that use Big Data and simulation technologies in the SCRM field (Kagermann et al., 2013; Lasi et al., 2014; Tiwari et al., 2018; A.A. Vieira et al., 2018; Zhong et al., 2016). In light of this, next subsection discusses the works related with the use of simulation to address SC problems, with a focus on the studies that use some type of data storage and integration structure to provide transactional real data to the simulation models. Thereafter, second subsection analyzes current research of structures to store, integrate and process data in Big Data contexts.

2.1. Simulation using data storage structures

According to Kagermann et al. (2013), the use of simulation to analyze the behavior of complex systems such as SCs should be emphasized. Simulation is even mentioned in one of the example applications provided by the authors, to analyze crisis scenarios in SCs. In fact, simulation has been extensively applied in SC problems, however, most studies use statistical distributions to model specific operations occurring in these networks (Jahangirian et al., 2010; Pires et al., 2016). For examples of such studies, see the studies of Cha-Ume and Chiadamrong (2018), Longo and Mirabelli (2008), Lee, Ghosh, and Ettl (2009), Chen, Mockus, Orcun, and Reklaitis (2012), Finke, Schmitt, and Singh (2010), Schmitt and Singh (2009), Blanco, Yang, Gralla, Godding, and Rodriguez (2011) and Mishra and Chan (2012). Conversely and according to Jahangirian et al. (2010), the absence of using real industrial data in simulation models may result in reduced stakeholders interest, with the cited authors noting that there is a gap of simulation studies making use of transactional data in simulation models. Exceptions for such cases are next analyzed.

Bottani (2008) reported the use of a Simul8 model representing logistic movements in a warehouse. It allowed to observe the impact of RFID (Radio Frequency Identification) in a warehouse, including the data that would be generated by such identification tags e.g., expiry date, type of component, production lot. Thereafter, this data was collected and stored in a DW, which, in its turn, was used to extract information from the simulation results. The authors emphasized the use of the DW structure to derive value-added information from the simulation results.

Cheng, Lee, Ding, Wang, and Stephens (2008) used GBSE (General Business Simulation Environment) to assess the performance of a SC comprised by 6 suppliers, 3 manufacturers and 3 customers, in order to quantify the transportation and holding costs, and the orders delivery time. Different production strategies (namely, make-to-order, make-to-stock and postponement), forecast accuracy levels and special transportation costs, were tested. All the data used in the model is stored in an external database (DB).

Schwede, Sieben, Song, Hellingrath, and Wagenitz (2009) developed a SC simulation model of the automotive industry using OTD-NET. A SC comprised by 3 manufacturers and 1 customer was considered, with the purpose of using the tool to help in entering in emergent markets. The results obtained from the simulation model were later stored in a DW, which provided further data analysis capabilities.

Advocating that typically used commercial tools present simulation results as averages that mask important aspects of the transient behavior of the system, Ehmke, Großhans, Mattfeld, and Smith (2011) aimed to improve the analysis of simulation results. Thus, the authors developed a DW that stored and allowed data analytics to be performed on the obtained simulation results. This way, it was possible to extract information from the simulation results. The simulation model was developed in PlantSimulation and the authors showed the benefits of such approach in a real case study, which considered transportation scheduling, operational procedures and infrastructural changes in the Mississippi River waterway system.

Nageshwaraniyer, Meng, Maghsoudi, Son, and Dessureault (2012) proposed an Arena simulation model of material handling operations of a coal mine real case study. In this study, the DW is used to feed data to the simulation model, while the results obtained by the later are also sent to the DW for analysis and further reporting. According to the authors, the use of both technologies allowed the best parameters for the system to be determined.

Rabe and Dross (2015) proposed a framework comprised by a DW and a simulation model, with the purpose of using simulation to predict changes that would occur to specific changes in specific logistics operations in a network of a company. In this framework, the main DW measured KPIs, while a copy of such DW was created to store simulation results. These simulation results, in their turn, were used as reward criteria for reinforcement learning algorithms, employed in the simulation model for accurate predictions.

Fornasiero, Macchion, and Vinelli (2015) and Macchion, Fornasiero, and Vinelli (2017) proposed a SIMIO simulation model, which assessed the impact of orders size in the SC performance. In (Fornasiero et al., 2015), the authors based their experiments on order size, lead time variation and supplier scrap rate to achieve findings regarding the impact on the delivery time to customer, customer order quality and inventory costs. Later, Macchion et al. (2017) assessed the impact of order size, inventory management policy, supplier's lead time and quality on inventory level and order lead time. Fornasiero et al. (2015) applied their model to a fashion industry comprised by 60 suppliers and 1 manufacturer, whilst Macchion et al. (2017) applied it to a SC of the footwear industry comprised by 4 suppliers, 1 warehouse, 1 manufacturer, 1 distributor and 2 customers. In both studies, the authors reported that their simulation models are able to retrieve data from the ERP system.

Sahoo and Mani (2016) presented a simulation model in ExtendSim to model a SC of the biomass industry. The modelled SC comprised producer and farmer of biomass and suppliers which transported the raw materials to the plant, which could store them or process them for later bioenergy production, in order to deliver heat and electricity to customers. The simulation model used a direct connection to a DB, which, among other operational data, stored weather data for long time periods. This data was used to test the exposure of the SC to KPIs (Key Performance Indicators) such as: quantity produced, inventory levels, transportation quantity, transportation costs and handling cost.

Lastly, Ponte, Sierra, de la Fuente, and Lozano (2017) evaluated the impact that inventory management and different forecast methods have on the demand variation propagation upstream the SC. The authors considered a linear SC with 4 stages: manufacturer, wholesaler, retailer and final customer. The simulation model also retrieves the required data from an external DB.

As the above reviewed studies suggest, the scope of most papers is reduced to a specific process, e.g. warehousing or transportation, failing to consider all the activities occurring in a SC. Moreover, the study of Fornasiero et al. (2015) presented the SC with highest size, with 60 suppliers providing material to a manufacturer. This seems to indicate that authors tend to select specific SC problems and select SC of reduced size or select specific sets of suppliers from the original SC, in order to reduce the complexity and, hence, the volume of data to be managed. In fact, all the reviewed papers considered traditional DBs or DWs, not using the benefits of Big Data concepts and technologies. Therefore, and to the best of the authors' knowledge, no study has used Big Data technologies to model SC systems, which is corroborated and whose importance and need has been widely recognized by several studies (Kagermann et al., 2013; Lasi et al., 2014; Tiwari et al., 2018; A. A. Vieira et al., 2018; Zhong et al., 2016). In addition, out of the reviewed studies, the work of Schwede et al. (2009) was the only one to considered the real case of an automotive electronics SC, which this paper also considers. For more examples of studies that considered automotive electronics SCs, see the works of Dev, Shankar, Dey, and Gunasekaran (2014) and Gansterer (2015).

2.2. Big data warehousing

SC systems generate data at increasingly higher rates, volumes and formats, in what is known as the three main characteristics of Big Data, which is a very recent concept (Zhong et al., 2016; Zikopoulos & Eaton, 2011). In fact, as asserted by Ularu, Puican, Apostu, and Velicanu (2012), despite being firstly used in the 1970's, only after 2008 the term Big Data started to gain considerable attention in literature. Due to the novelty of this research filed, it still does not have a widely accepted definition, which may be dependent on how technologies will evolve, as suggested by Costa and Santos (2017a). In light of this, it is hard to quantify a well-accepted threshold to a Big Data context. Therefore, many times, this value is defined as the volume that exceeds the capacity of traditional tools to process the data (Kaisler, Armour, Espinosa, & Money, 2013; Madden, 2012).

The traditional approach for data storage and analysis consists in using Online Analytical Processing (OLAP) systems, such as Data Warehouses (DWs). However, traditional relational data models are less effective and efficient to work in the described Big Data contexts, which paved the way for a new solutions to continue providing organizations with analytical and decision-support capabilities (Costa & Santos, 2017b; Costa, Costa, & Santos, 2017). In an attempt to propose other approaches for Big Data contexts, different types of solutions have been proposed and implemented. As Costa et al. (2017) and Costa and Santos (2017b) suggested, some considered implementing DWs in NoSQL DBs, albeit these solutions are only scaling operational systems (see Lourenço, Abramova, Vieira, Cabral, & Bernardino, 2015) for a comparison of NoSQL engines). Eventually, SQL on-Hadoop emerged as a more efficient solution for Big Data environments (C Costa & Santos, 2017b; E. Costa et al., 2017; Grover & Kar, 2017; Mohanty, Jagadeesh, & Srivatsa, 2013). See (Goss & Veeramuthu, 2013) for a comparison of Hadoop and other alternative solutions for Big Data contexts and Grover and Kar (2017) for a summary of existing Big Data tools, including the Hadoop ecosystem.

Hadoop is an ecosystem based on the MapReduce programming model and the Hadoop Distributed File System (HDFS). Several systems are included in it, such as Hive, Impala and others, which are used for different tasks required under Big Data contexts. Hive is widely adopted by many organizations and was created by Facebook as a way to improve the Hadoop query capabilities that were very limiting and not very productive (Thusoo, Sarma, et al., 2010). At Facebook, it is extensively used for reporting, ad hoc querying and analysis (Thusoo, Shao, et al., 2010). Hive organizes the data in tables (each table corresponding to an HDFS directory), partitions (sub-directories of the table directory) and buckets (segments of files in HDFS). In addition, it has its own query language: the HiveQL (Hive Query Language). Thus, a DW developed in Hive can be seen as a BDW, being a flexible, scalable and highly performant system that uses Big Data techniques and technologies to support mixed and complex analytical workloads (e.g., streaming analysis, ad hoc querying, data visualization, data mining and simulations) (Carlos Costa & Santos, 2018).

Santos et al. (2017) presented a Big Data architecture implemented in Bosch Car Multimedia in Braga, Portugal (the same plant of the case study considered in this paper), which supports the Industry 4.0 technological movement followed by the organization in question. The developed Big Data system integrates data from several business processes, like customer quality claims, allowing the analysis of several Key Performance Indicators (KPIs) in this area and was implemented in the Hadoop ecosystem. Nodarakis, Sioutas, Tsakalidis, and Tzimas (2016) extracted hashtags from large scale tweets to classify them into different sentiments, in a parallel and distributed manner, using the same ecosystem. The authors also conducted experimental evaluations to demonstrate that their solution is efficient, robust and scalable, hence, being appropriate for Big Data contexts. Kv and Kavya (2016) also used Hadoop to analyze trends of e-commerce web traffic logs.

3. Methodology

This section discusses the methodology applied for this research. Thus, first subsection describes the case study adopted for this research and next subsection discusses the adopted approach.

3.1. Automotive supply chain

This project is being developed at a plant of the Bosch Group, which concerns with producing electronic car components. Car manufacturers need to comply with very strict security norms for their products, while still providing high levels of product customization, required by increasingly demanding end customers (Masoud & Mason, 2016; D Simchi-Levi et al., 2015). At the same time, an ordinary car is comprised of roughly 5000 parts and materials are usually supplied by single sources, exposing manufacturers to specific suppliers, thereby posing a disruption danger for the entire SC (Thun & Hoenig, 2011). Hence, entities interoperating in these SCs need to comply between them, in order not to jeopardize the entire chain (Kırılmaz & Erol, 2017).

To streamline SC processes, the automotive industry has incorporated concepts such as Just in Time (JIT), a pull philosophy concerned with demand-driven production, aiming to reduce overall wastes, namely with inventory levels (Masoud & Mason, 2016; Thun & Hoenig, 2011). Indeed, the adoption of these concepts results in low inventories on one hand, but high vulnerability on the other (Ghadge, Dani, & Kalawsky, 2012), because available materials may not be sufficient to cover eventual disruptions. To deal with these eventual failures, entities in these SCs tend to adopt safety stock approaches, to ensure material buffers in case of disruptions. However, establishing proper safety stock levels is also a complex task, because, on one side, low safety stocks can lead to eventual stockouts, whilst the opposite can result in overstock, leading to further unnecessary warehouse costs.

The plant considered in this case study is located in Portugal and has suppliers and customers from all around the world. To provide a perspective of the scale and complexity of the network in analysis, Fig. 1 shows the countries and the number of suppliers per country, which supply materials to the plant.

The numbers assigned to each country in the above figure represent the number of suppliers from each country. The lines placed between these countries and the Braga plant represent the number of material shipments; their width is proportional to the number of shipments. Finally, the color scale of each country is associated to the number of different types of materials that suppliers from that country provide to the plant. According to the data used to create Fig. 1, around 7 000 different types of materials are actively being supplied by roughly 500 different suppliers, located in more than 30 countries. Moreover, as the figure shows, Germany, Netherlands, Switzerland, Spain, China, Taiwan and Malasya are the countries that supply more types of materials. Also suggested by the above figure, the Braga plant received materials from more than 400 suppliers, especially from Europe and Asia, with Germany (209 suppliers) and Netherlands (10 suppliers) having more suppliers and shipments from Europe, and Malasya (16 suppliers), Taiwan (13 suppliers), China (12 suppliers), Hong Kong (11 suppliers) and Singapore (7 suppliers) having more shipments from Asia.

3.2. Proposed approach

This section describes the methods employed in this research, which comprise the development of a Logistics BDW that stores, integrates and provides real industrial data to an automotive SC simulation model. In this regard, the Big Data cluster of the Bosch organization was used and Fig. 2 depicts the framework defined for this research.

The first step of the research consisted in the identification of the relevant SC processes. For this purpose, several interviews, focus groups and other group sessions with process experts were conducted. Next, it was also necessary to select the associated data sources. Notwithstanding, it should be noted that in big organizations, the same business process may be managed by different systems, hence, being necessary to analyze, with process and data experts, which data sources should be considered. See (Vieira et al., 2018) for further details of this step. This analysis culminated in the inclusion of the following sets of business processes and its corresponding data sources:

- Internal Movements – Managed by a SAP transaction;
- Orders to suppliers – Managed by a SAP transaction;
- Suppliers' data – Managed by SAP and using an additional Excel file;
- Materials' master data – Managed by multiple SAP transactions;
- Material shipments or On Time Arrival report – Managed by an Excel file and later inserted in SAP;
- Early arrivals – Managed by an Access DB;
- Special freights – Managed by an Information System developed at the plant.

Despite the above list, there are other business processes that should be considered by a SC simulation model aiming to mimic the behavior of the real system, but could not be considered by this research, due to issues that were found related with the associate data sources, e.g., data source does not cover the intended time frame, data source does not exist, no access to sensitive data. In fact, the multiple problems that can be found when using real industrial data have already been analyzed by previous research, e.g. see the works of Bokrantz, Skoogh, Lämkull, Hanna, and Perera (2018) and Wang and Strong (1996). As such, the above list corresponds to the business processes (and the respective data sources) that could be obtained. Thus, other business processes, such as final customers' orders, orders' forecast, multi-tier suppliers' data should also be considered.

To collect data from these sources, the ETL approach was followed, meaning that data is collected, the necessary transformations are performed, in order to correct eventual data issues and, thereafter, data is sent to the staging area, namely to the HDFS system. The Talend

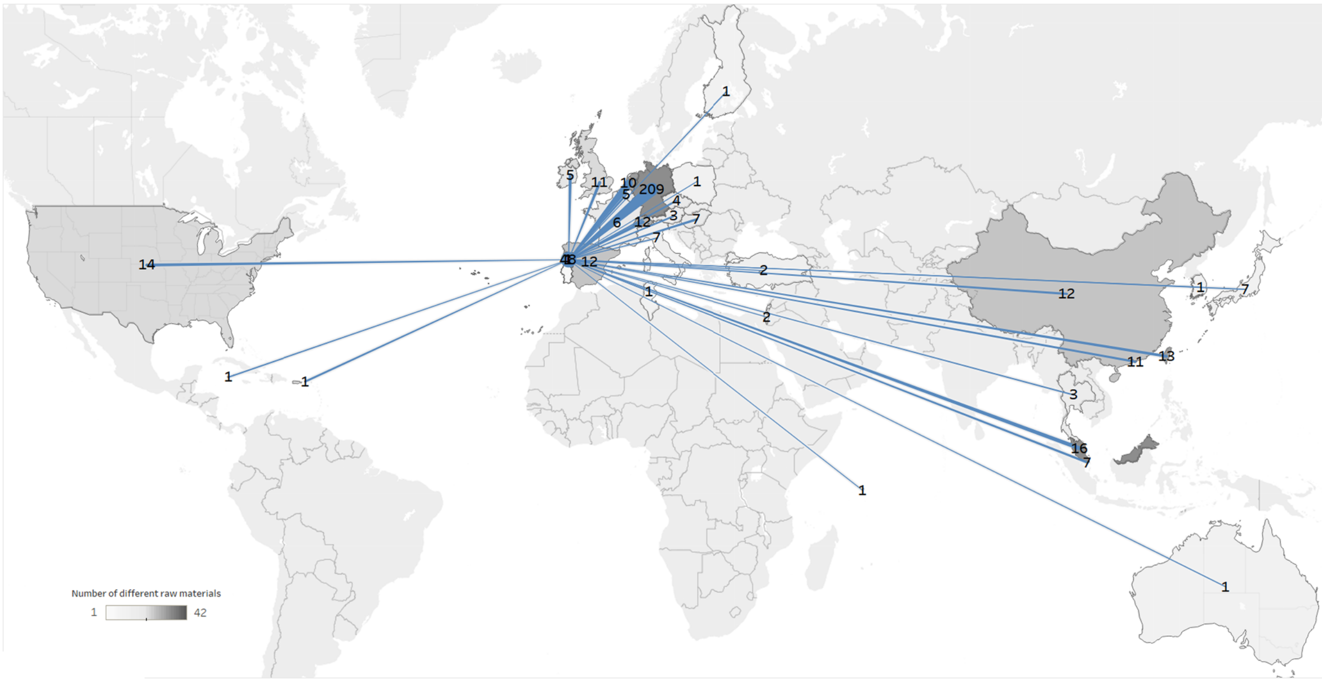

Fig. 1. Geographic location, number of suppliers, different raw materials provided and number of shipments per suppliers.

software was used for this task and additional details can be seen in (Vieira et al., 2018). The same software was used to create the BDW tables (in Hive), in the Hadoop ecosystem of the Big Data cluster (see Section 2.2 for additional details regarding the Hadoop ecosystem).

In traditional data modeling approaches (e.g., DWs), data is usually stored in multiple data tables, where facts (typically quantifiable values) are stored in fact tables and the remaining descriptive values are stored in dimension tables. This approach, on one hand, reduces the space occupied. However, on the other hand and particularly in simulation contexts, it increases the computational resources required to run simulations in which entities need to search for data among multiple tables with considerable size both in number of rows and columns per table. Such problem is even heightened in Big Data environments, where the volumes of data are even higher and generated at increasingly higher rates. Therefore, in this case, the solution consisted in reducing the number of Hive table and modeling the data according to the data needs of the simulation model. Hence, the data in these tables is completely denormalized, so that the simulation engine does not need to search for any value during runtime. For instance, a single Hive table stores all data related with the orders placed to suppliers and the respective arrival.

4. Simulation in a big data context

In this section, the modeling approach that was conducted to develop the simulation model that runs in a Big Data context will be discussed. In this regard, first subsection focuses on the modeling approach. Next, second subsection discusses the volume of data that was managed in this research, i.e., for a first instance of the implemented Logistics BDW.

4.1. Simulation model

This subsection presents the main modeling steps conducted to develop the SC simulation model. As stated in Section 3.2, the BDW was modelled in accordance to the simulation needs and the data per Hive table was fully denormalized. Therefore, when an entity is created, a representative entity is also created and associated to the same data row and the entity keeps the data as its attributes and activities to perform in the simulation model. To model the behavior of the entities, several processes were developed, allowing entities to freely move in a 3D environment. Fig. 3 shows two examples of processes that were created. The one on the top models the act of placing orders to suppliers and the one on the bottom models their production and shipment back to the plant. Note that these figures do not provide the complete view of the processes, as it was necessary to simplify them, with the purpose of explaining the approach, i.e., the real ones are considerably larger. Furthermore, many more processes were developed to model the behavior of entities in the remaining situations of the system, e.g. internal material movements and production.

The process at the top of Fig. 3 starts by doing some initial assignments to an entity, which consists of reading values from the respective data row of the Hive table with all orders to suppliers. Thereafter, the entity is sent to the supplier location. This activity represents the act of placing an order to a supplier, which was modeled so that users could visualize the orders being placed to the respective suppliers, being a requirement from stakeholders. In addition, this was modeled at very high speeds and its duration was considered when obtaining the respective order lead time. Finally, to achieve this, entities enter in Free Space movement (using the Transfer step), which is a feature offered by SIMIO that allows entities to travel between objects without requiring any link or connection between them. Thus, the movement of entities needs to be modeled using Travel steps, which adds complexity to the development and allows entities to freely move in an orthogonal tridimensional space, as connections between two distinct locations are not used. Moreover, this feature also enables the automatic update of the simulation model, as long as these changes are reflected in the data, e.g., if new suppliers are added or if materials start being provided by other suppliers, as long as these changes are reflected in the BDW. In other words, since there are no physical non-entity objects in this model, including links, the simulation model is generated by the data and the processes model the behavior of entities, in a data-driven approach.

While at the respective suppliers, the entities verify if delays occur. If no delays occur, the transit and lead times are obtained from the BDW and it is verified if the respective order arrives in due date, earlier than scheduled or if it will arrive later than scheduled. Thus, while early arrivals may result in higher warehousing costs for a plant, late arrivals

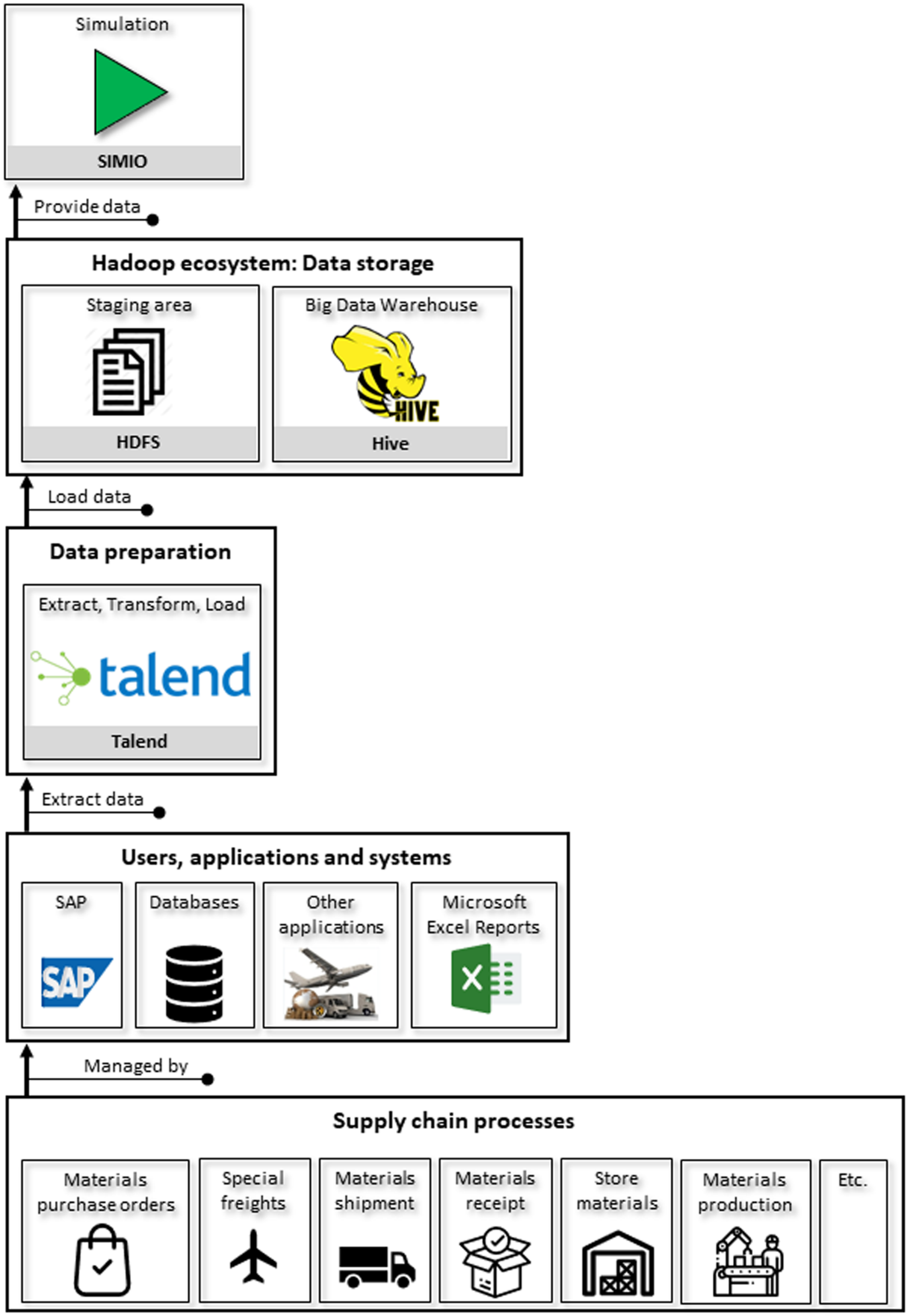

Fig. 2. Framework used to develop the BDW and the simulation model.

may originate production stoppages as the plant follows a JIT philosophy, therefore, in the latter situations, special freights need to be scheduled, which have considerably higher costs. Finally, the process verifies if the supplier in question is within an area affected by a geographic disruption. If not, then the process represented at the bottom of Fig. 3 is executed, which models the production of orders, using the "Lead time" Delay step, and consequent transportation back to the plant, using the "Go to plant" Travel step. Finally, the process verifies if other delays exist and, afterwards, transfers the entity to the plant, removing it from Free Space.

As the processes illustrate, several disruptions were included in the model. Thus, the simulation model is able to mimic the behaviour represented by the data stored in the BDW and include uncertainty through the inclusion of disruptions or variability. However, in the case provided in this paper, only geographical disruptions are considered. I.e., disruptions that occur in a given location and affect all suppliers in

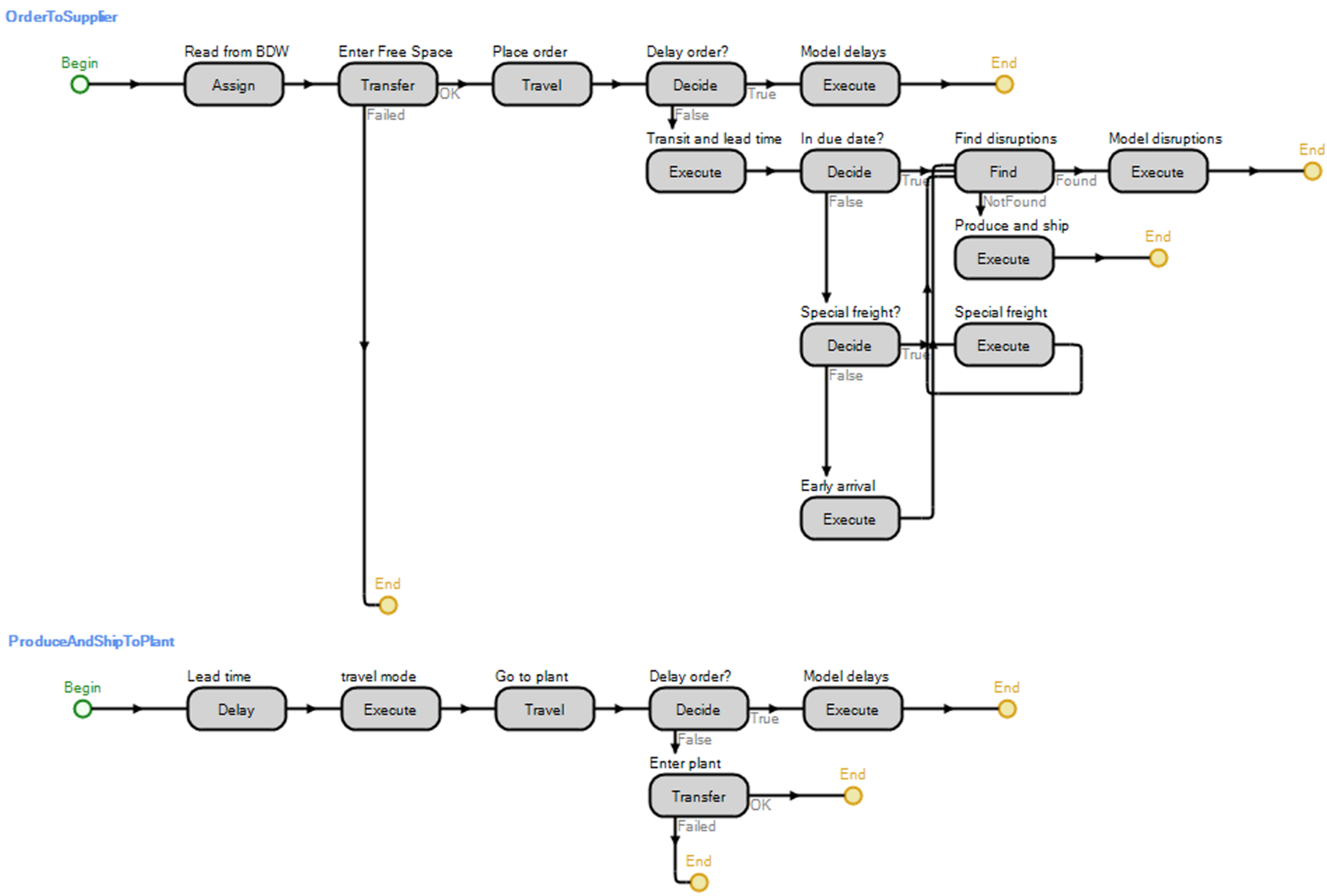

Fig. 3. Processes used to model an order being placed and its arrival to the plant.

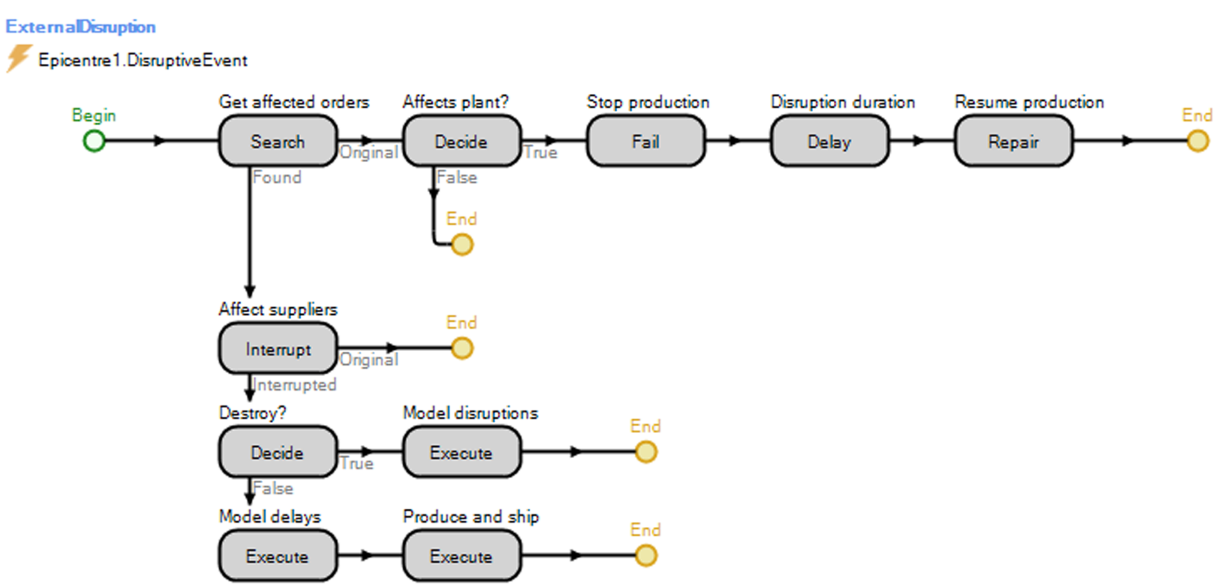

Fig. 4. Process used to model a disruptive event in the simulation.

that location, which can be triggered by natural disasters, labor strikes, political instability and others. Fig. 4 shows the process created to model these disruptive events.

The process is triggered by an event (Epicentre1.DisruptiveEvent), which, in its turn, may be fired at any simulation time, with any duration and location, with any range and any number of times. When the execution of the process is triggered, it first searches in the tridimensional world map for all entities that represent orders located within a certain specified range. The entities found by this search may be destroyed or delayed, depending on the simulation parameters. Thus, this process is able to interrupt other entities representing orders that are being processed by suppliers. After searching for the affected orders, it evaluates if the plant itself is located within the specified range. If true, then the plant's production is halted during the specified disruptive duration, resuming it afterwards.

This subsection provided some aspects regarding the development

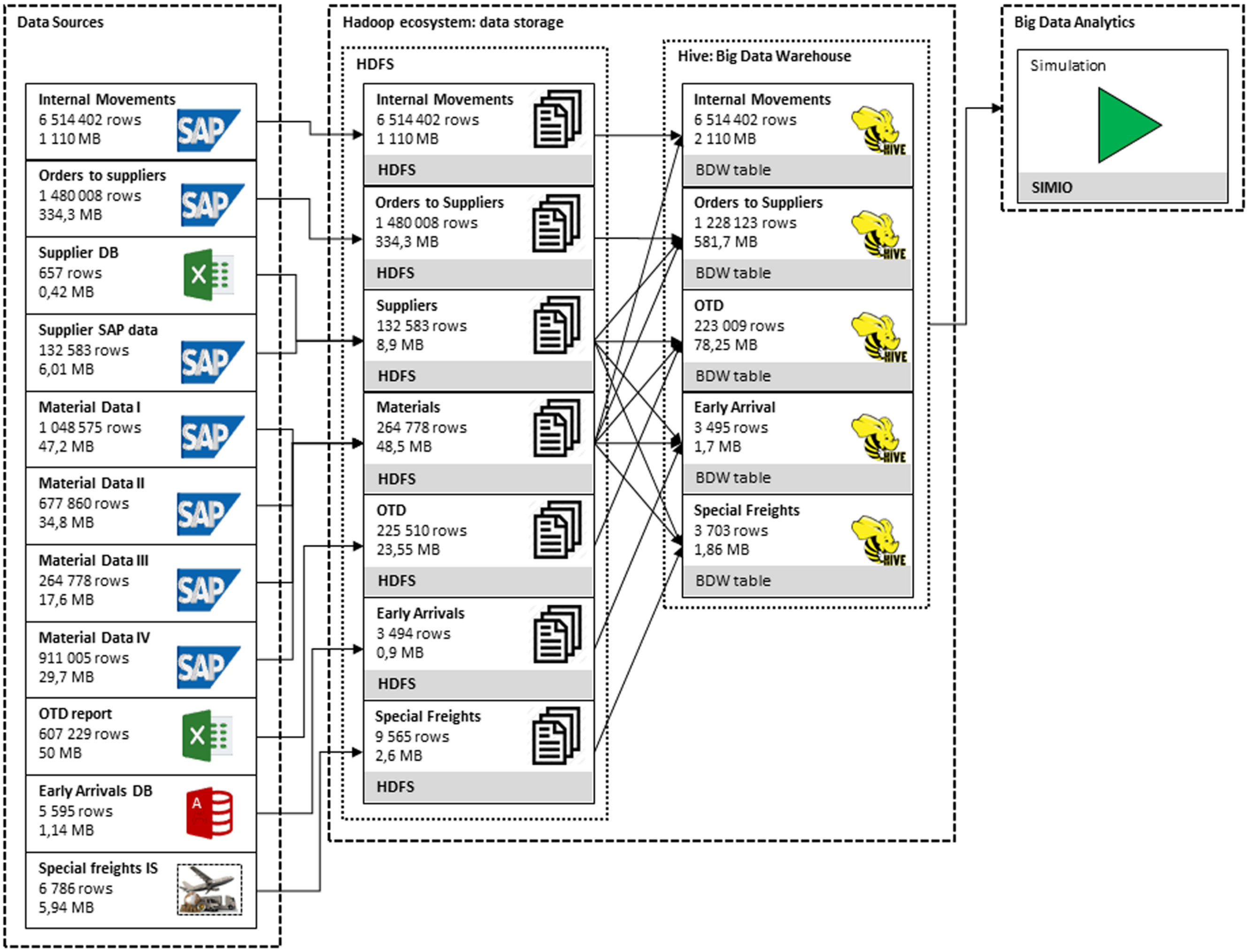

Fig. 5. Summary of the volume of data managed in the ETL process.

of the simulation model with emphasis on the approach that allows the simulation model to mimic the behavior of the system and also include disruptions that did not occur, allowing alternative scenarios to be tested. Next subsection discusses the volume of data that was managed in this research.

4.2. Supply chain Big Data Warehouse: a first instance

This subsection analyzes the volume of data handled in this research, serving as a milestone for future researches in the same domain. In this regard, Fig. 5 shows a summary of the volume of data, which considered 1 year.

On the left side of the figure, the volume of data that was obtained from each data source can be seen. In the middle, the Hadoop ecosystem is represented, which includes the data staging area (HDFS) and the Hive tables. As can be seen, data from the data sources is combined in an HDFS file for each business process involved, i.e., internal movements, orders to suppliers, suppliers' data, materials' data, OTD report, early arrivals and special freights. For instance, several SAP transactions contain relevant data for materials, thus, all data is merged in a single HDFS file. Next, data in the HDFS is combined to create the Hive tables, according to the simulation needs.

It can be seen that the internal movements (SAP) provided the highest volume of data, with roughly 1 GB (gigabyte) of data. After denormalizing the data with the associated data of materials, a Hive table with more than 2 GBs was obtained. Hence, the BDW provided roughly 3 GBs of data to the simulation model. Another interesting aspect is related with some of the differences between the volume of data extracted from the data sources and the volume sent to HDFS. This is related with the operations that were done in the ETL process, i.e., whilst some operations required filtering data, others required duplicating rows, e.g., when several materials are specified in the same data row of the data source.

By solely analyzing the volume of data, it is indeed arguable if this environment can be considered a Big Data one. Notwithstanding, as Madden (2012) and Kaisler et al. (2013) stated, there is no widely accepted threshold value for a Big Data context. Thus, often, this value is defined as the volume that exceeds the capacity of traditional tools to process the data, hence being necessary to apply Big Data concepts and technologies. Such was the case with this research, where the organizations' Big Data cluster was used and Big Data concepts (namely by denormalizing the data) and technologies (namely the Hadoop ecosystem and the therein included Big Data tools) were applied.

Furthermore, this is a first instance of the BDW, meaning that it must be continuously updated, so that updated data can be provided to the simulation model, which implies that the volume of data will exponentially increase, as per a Big Data context. In addition, this volume of data corresponds to a single year and to the one that could be used in the project. I.e., as discussed in Section 3.2, some data or even some data sources could not be considered, thus, it would be expected that with this additional data, the volume in the BDW would further increase, heightening the importance of applying the Big Data tools and concepts addressed in this research.

5. Results

This section shows the main results that can be obtained from the developed simulation model that runs in a Big Data context. Thus, in the first subsection, the animation of the model is analyzed. Next, in the second subsection, the simulation model is used to test the performance of the system, when suppliers in a geographic area are disrupted.

5.1. Animation

The benefits of animation in simulation are widely known and recognized, especially in the context of Industry 4.0 (Brettel, Friederichsen, Keller, & Rosenberg, 2014; Posada et al., 2015; Turner, Hutabarat, Oyekan, & Tiwari, 2016). In this case, the model was developed in a tridimensional environment and using an integration with Google Earth. Such features improve the quality of the animation of the model as will be illustrated in this subsection. In fact, such enhanced quality was a requirement of the stakeholders from the plant that were involved in the project.

As the simulation clock advances, when it reaches the date of a specific order to a supplier, an entity is created, which represents that order. As mentioned in Section 4, this movement is depicted in the model, to illustrate orders being placed to the respective suppliers. Fig. 6 illustrates this situation.

As can be seen, the model runs in a world map view. The blue triangle-shaped entities represent entities being placed to suppliers. The name of the countries represented below each entity represents the country where the supplier is located. The figure also shows some circles placed north of Portugal. The yellow ones represent orders that have already been placed to the respective suppliers and, as such, are being processed. It should be noted that the location of these entities represents the exact location of the supplier, as stored in the BDW. However, as a supplier may have multiple orders at the same time, a small deviation in the location of each order is applied, so that it is possible to see all entities. Finally, the number presented above each yellow entity represents the number of days remaining for the order to be shipped to the Braga plant, i.e., the remaining lead time, which decreases as the simulation clock advances in time. When it is time to ship the order, the symbol of the orders changes to the respective transportation type and their speed is also adjusted to the transportation duration, as stored in the BDW. Fig. 7 shows some of these entities highlighted.

The figure shows one aircraft delivery and groups of truck deliveries highlighted with dashed circles. Yet, albeit not illustrated in this figure, sea craft transportations also exist in the real system. The date time values below each entity represent the instant when those deliveries were shipped to the Braga plant. When orders arrive to the plant, they are managed by the data source of the internal movements, which includes the production orders.

The animation of the model allows to visualize both material and information flow that occurred and estimated alternative ones, hence allowing to discover new patterns in the huge amounts of data stored in the BDW. The screenshots provided in this paper testify the quality of the animation of the simulation model. Next subsection concerns with the analysis of the results obtained for the simulation experiments conducted with the proposed tool.

5.2. Simulation experiments with geographic disruptions

This subsection analyzes some simulation experiments which focus on assessing the impact of geographic disruptions in the performance of the system. When these disruptive events occur, an area is affected for a certain duration, leading to consequences hard to properly assess. Thus, first, it shows how such disruptions can be included in the simulation

Fig. 6. Simulation model when orders are being placed to suppliers (North of the Iberian Peninsula).

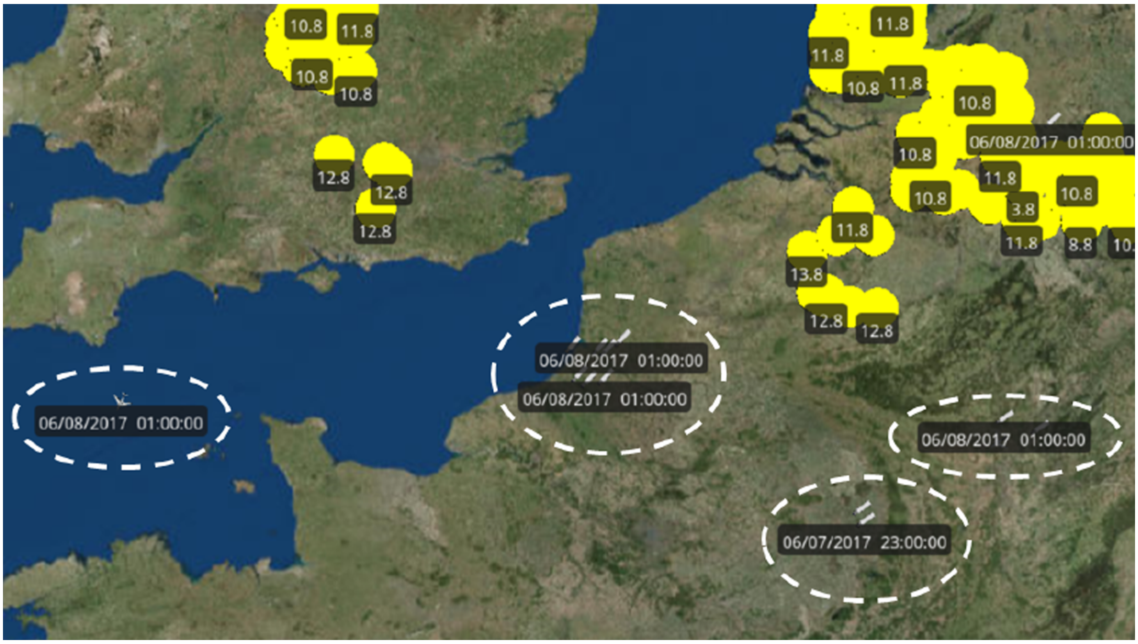

Fig. 7. Orders being shipped to the Bosch Braga plant (North of Europe).

Fig. 8. Disruption event being triggered during simulation run time.

and, thereafter, its impact in the performance of the system is analyzed. These events are incorporated in an interactive way, in what is known as visual interactive simulation (Robinson, 2005), meaning that the user may fire several events and choose, in runtime, when to trigger the event, which geographic location to affect, the diameter range and its disruption duration. Fig. 8 shows the moment when this button is clicked.

The figure shows a button being clicked. At this instant, an event is fired, which triggers the execution of the process depicted and already described in Fig. 4. The distance between the button and the *"Limit@DisruptiveEvent1"* node defines the range of the event. Thus, all orders inside this perimeter are affected by the disruption event, during the defined disruption time (defined in the simulation parameters). As can be seen, the orders change their color to red, when they are disrupted.

To demonstrate the results that can be obtained with simulation, the following two Key Performance Indicators (KPIs) were chosen: the average stock percentual difference and the number of unfilled orders. Furthermore, three scenarios, with a simulation time of one year, were

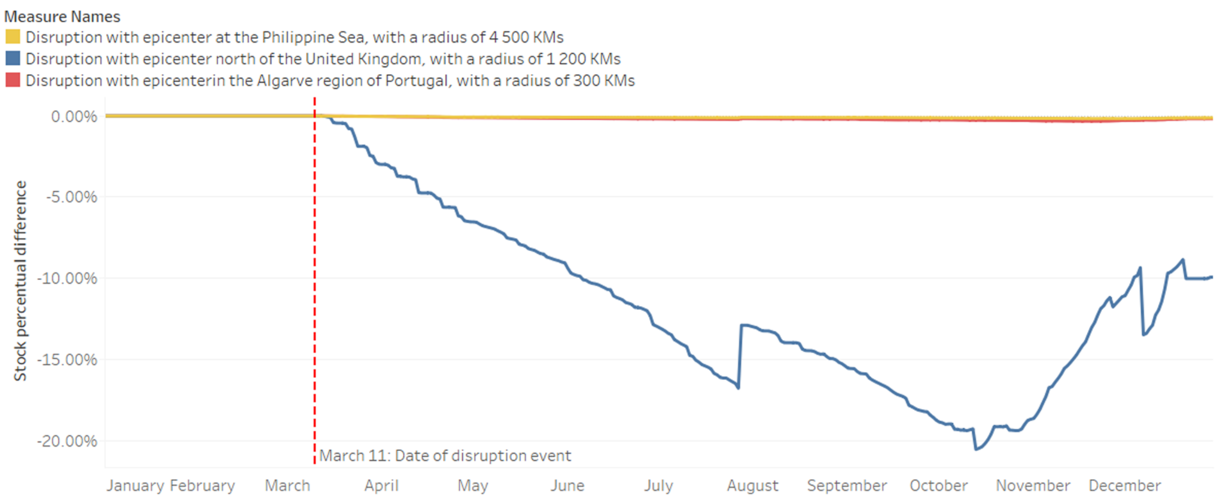

Fig. 9. Stock percentual difference for the 3 scenarios.

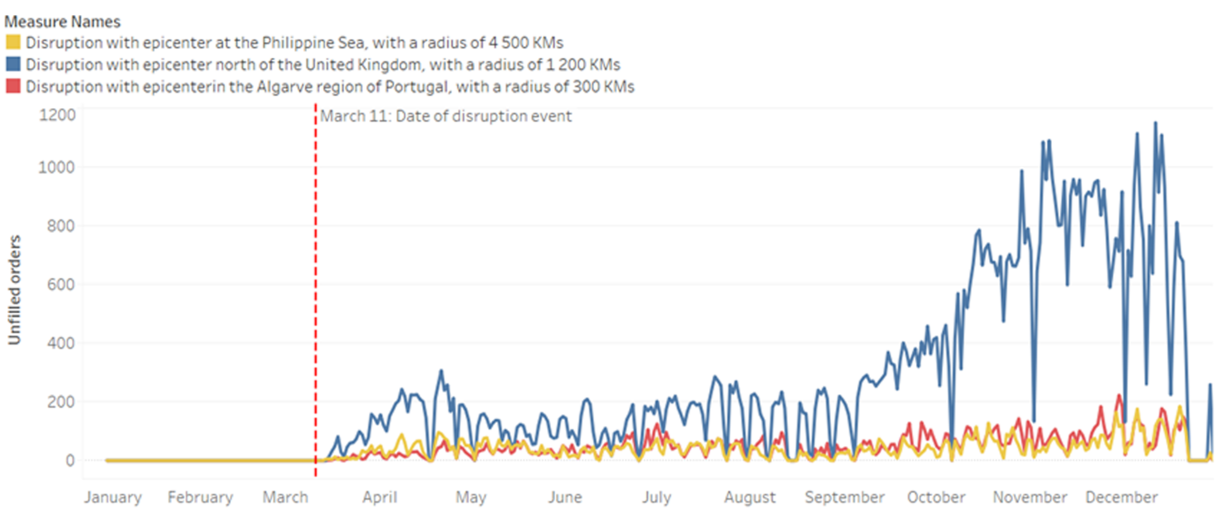

Fig. 10. Number of unfilled orders for the 3 scenarios.

considered. The first scenario reproduced a simulation from the historical data stored in the BDW. In its turn, the remaining two scenarios also reproduced the historical data of the BDW, however, in these scenarios, an event which disrupts the operations of suppliers located in a certain geographic area was triggered. The purpose of these events is to destroy the entities regarding the orders that are in lead time at the respective suppliers and assess the corresponding impact. Second and third scenarios considered different geographic zones and different ranges of radius, as indicated in Figs. 9 and 10,which show the results obtained for the defined KPIs.

As the figures show, even though the disruptive events fired in the considered scenarios only impacted the suppliers in a single day, the corresponding consequences are felt throughout the year. This can be explained by the type of industry, i.e. automotive, in which most plants are supplied by single sources, therefore, when such suppliers fail to deliver the orders considerable impacts may be felt, as reported in related research (Matsuo, 2015; Park et al., 2013).

The consequences are also felt with different magnitudes among the considered scenarios. In fact, even though the disrupting event fired at Asia impacted a wider area than the remaining locations, the highest impact was felt when the disruption occurred in the north of Europe, revealing a greater dependency on suppliers from this region of the world. The consequences of the disruptions are felt only roughly 3 days after the disruption event occurs, which is justified by the JIT philosophy followed at the plant, i.e., even though the stocked materials have safety stock, those buffers are short, in order to avoid high stock levels.

Despite the high impacts to the SC system, that the considered scenarios provided, it corresponds to situations in which reactive measures were not modeled. In other words, when facing disruptions such as the ones considered in this subsection, the plant should try to mitigate them using several strategies, e.g., by contacting alternative suppliers, or even by verifying the possibility of alternative production plans. However, to be able to do this, more data would be required, e.g., data related with situations similar to the ones modeled in this subsection. Not all this set of additional data could be obtained, thus, the results presented in this subsection correspond to the highest possible impact.

The results discussed in this section indicate the benefits of the proposed tool for decision-makers, which are aligned with the expected insights that can be obtained from a simulation solution, e.g. visualize complex systems and test alternative scenarios, albeit in this subsection the simulation model was only used to test scenarios that considered disruptions fired by users during runtime. Notwithstanding, the tool can also be used for other purposes, e.g., test several types of variability. However, this set of features is limited by the lack of data already discussed. For instance, to properly use a simulation model for prediction, it would be necessary to apply distribution fitting and seasonality methods and have access to considerable amounts of data to apply such methods, which would confer the adequate behavior to the model.

Nevertheless, this subsection has provided an example of a way in which a simulation model, running in a Big Data context, can be used to enhance decision making in SCRM.

6. Managerial implications

This research considered a real case study of a manufacturer that is part of an automotive electronics SC and has its own business processes and data sources. Such elements must be considered when developing the respective simulation model, as they must mimic the behavior of the business processes using the available data. The same applies to the BDW, as it will provide a data repository that stores data originated from the data sources used at the plant. In fact, even plants belonging to organizations with strict standardized norms have their own tailored software and their business processes that do not apply to other plants of the same organization; such was the case with this plant. The result is a BDW and a consequent simulation model that are tailored for the plant of the case study. Nevertheless, this research provided some insights, from a managerial perspective, that are discussed in this section.

In this regard, the improvement in the decision-making, in multiple perspectives, should be emphasized. First, typical reporting in Logistics Departments of organizations consists of static reports. In fact, the same type of reporting was used at the plant hosting this research. Thus, with a BDW, since it integrates data from several sources and applies Big Data concepts and technologies, it can be used for several purposes, e.g., feed data to simulations, ad-hoc querying, dynamic reports with interactive dashboards, Artificial Intelligence algorithms and others.

Second, the simulation model allows to adopt a proactive approach in SCM. In fact, in SCs with considerable size and complexity, such as typical automotive SCs, it is often hard to properly quantify the impact of certain disruptions. Thus, as demonstrated in this paper, the simulation model allows to test disruption scenarios of several types, albeit in this paper only geographical disruptions were covered, which were fired in an interactive way. Such approach may be interesting for managers, as this type of analysis allows to observe the reaction of the system, by testing different parameters, e.g. disruptions fired at the beginning or the end of simulation time, at Asia or at Europe. In fact, such type of analysis was required by the organization hosting this research, as an insightful type of analysis that the simulation tool should allow. Furthermore, as discussed in the paper, the simulation model can also be used for other purposes, e.g., prediction. However, to allow some of these features to be used, possible problems with data may need to be managed.

Third, the authors argue that proper SC simulation models benefit from the use of Big Data concepts and technologies, as they allow the detail of all related business processes to be captured, which traditional simulation approaches (using statistical distributions) fails to do with the same level of efficiency. While traditional structures such as DWs also allow to provide real data to simulation models, the volume of data that this type of systems generate and the need to avoid simulations to search data among different tables during runtime (otherwise simulations can become considerably slow), strongly justifies the need to use Big Data concepts and technologies in SC simulations.

Fourth, the approach followed in this paper differs from traditional simulation studies in that it does not use statistical distributions to model the business processes of the considered system. Rather, it uses the data stored in the BDW, thus, being able to reproduce an exact copy of the real system. On one hand, this approach is especially demanding as it requires the developer to be closely involved with data and process experts, in order to identify the data sources that should be used for each relevant business process. On the other hand, it fosters the engagement of all stakeholders in the project and increases their interest and involvement in the tool. The need to use real data in the project may also result in finding data problems (Bokrantz et al., 2018), further heightening the need to and the importance of the engagement with stakeholders.

Finally, and having been the first research to discuss the development approach of a SC simulation model that runs in a Big Data context, future researches and practitioners working in SCM may also benefit from this work with the methodology and the Big Data concepts and technologies that were applied and shared in this paper, as well as withdrawn conclusions and insights.

7. Conclusions

SCs are known to be complex and dynamic networks, which are prone to uncertain events that may affect its performance. Thus, proper DSS are required, in order to mitigate the impact of such risks, allowing proactive measures to be taken. For this reason, simulation may be used, as it allows alternative scenarios to be tested, among other benefits. In addition, in today's Industry 4.0 age, SC systems generate data at increasingly higher rates, volumes and formats, hence creating a Big Data environment. In light of this, this paper considered the real case of an automotive electronics SC and proposed a DSS comprised of a BDW and a simulation model. The former stores, integrates and provides data to a SC simulation model, which can be used to mimic the behavior of the system and test alternative scenarios, e.g.: to quantify the impact of disruptions in the performance of the system.

The unique contributions of this paper to the state-of-the-art can be considered as a SC simulation model that runs in a Big Data context, being able to mimic the behavior of the real system and, at the same time, allowing risks to be analyzed. Hence, the first step of the project was to allow the simulation model to mimic the behavior of the system, according to the data stored in the BDW, which opens room for using the tool for other tasks common for simulation studies, e.g., test risk scenarios. While this paper only considered risks that occur in geographic areas, other types of risks can also be analyzed with the proposed tool. While it is, indeed, arguable if the volume of data that was used in this research can be considered to be Big Data, the fact that Big Data technologies (e.g., Hadoop ecosystem) and concepts (e.g., fully denormalizing data) were applied, strongly supports the claim that this research considered a Big Data environment. In addition, being a first instance of the BDW, it is expected that the volume of data stored in such structure will continue to exponentially increase, further heightening the importance of the approach followed in this paper.

Despite this Big Data context, several issues with data were found. In fact, certain disruptive events may also have indirect impacts in the plant. For instance, the disruption of a supplier could result in failures in other suppliers located outside the disruption zone, which, in its turn, could lead to a higher impact to the manufacturing plant considered in this case study. This issue can only be solved with complete SC integration and availability of the data of all the members of the SC, as emphasized by Levi et al. (2008).

All the mentioned problems with data occurred in an organization known for its excellence in technology and innovation and whose business processes are a reference in the automotive electronics industry. Notwithstanding, this leads the authors to conclude that, despite the current trends emphasizing the need to couple SC simulations with Big Data technologies, organizations are still struggling with the quality and availability of data that would allow accurate mimics of their SC systems to be achieved. This is especially relevant in simulation studies, as these will produce dynamic models based on the data they get. Thus, if relevant data is missing, the simulations may lose reliability. Nevertheless, it is expected that such issues may be somewhat mitigated as the Industry 4.0 completely materializes, allowing data to be automatically collected, treated, stored, integrated and provided to simulation models.

Regarding future research directions, SCs are known to be quite dynamic, i.e., the relations between agents may frequently change. Such changes may result in the simulation model and the BDW to be no longer accurate. As such, efforts must be made towards allowing both to operate in real-time. Regarding the former, it should automatically adapt to the data stored in BDW, even if data changes occur, using, for instance, data-driven approaches. Regarding the BDW, other Big Data concepts can be applied to allow this real-time feature. In this regard, the real-time interoperability between systems exchanging data must be ensured, e.g. between SAP and the BDW.

Finally, with the advent of artificial intelligence methods, such as machine learning or deep learning, it seems reasonable to assert that the next window of opportunity that simulation practitioners need to capitalize on concerns with the use of these technologies in SC simulation models in Big Data contexts. Hence, using the available data and Artificial Intelligent algorithms to confer actual intelligence to the agents considered in simulations, would allow simulation models to consider not only the individual behavior of agents, but also to observe their actions when they gain "intelligence". Such approaches could complement the tool proposed in this paper, including other approaches, e.g., optimization (Rocha de Paula, Boland, Ernst, Mendes, & Savelsbergh, 2019), selective maintenance (Duan, Deng, Gharaei, Wu, & Wang, 2018) and others.

Acknowledgments

This work has been supported by FCT (Fundação para a Ciência e Tecnologia) within the Project Scope: UID/CEC/00319/2019 and by the Doctoral scholarship PDE/BDE/114566/2016 funded by FCT, the Portuguese Ministry of Science, Technology and Higher Education, through national funds, and co-financed by the European Social Fund (ESF) through the Operational Programme for Human Capital (POCH).

References

Azzi, R., Chamoun, R. K., & Sokhn, M. (2019). The power of a blockchain-based supply chain. *Computers & Industrial Engineering, 135*, 582–592. https://doi.org/10.1016/j.cie.2019.06.042.

Blanco, E. E., Yang, X., Gralla, E., Godding, G., & Rodriguez, E. (2011). Using Discrete-event Simulation for Evaluating Non-linear Supply Chain Phenomena. In Proceedings of the Winter Simulation Conference (pp. 2260–2272). Winter Simulation Conference.

Bokrantz, J., Skoogh, A., Lämkull, D., Hanna, A., & Perera, T. (2018). Data quality problems in discrete event simulation of manufacturing operations. *Simulation, 94*(11), 1009–1025. https://doi.org/10.1177/0037549717742954.

Bottani, E. (2008). Reengineering, simulation and data analysis of an RFID system. *Journal of Theoretical and Applied Electronic Commerce Research, 3*(1), 13–29.

Bottani, Eleonora, Murino, T., Schiavo, M., & Akkerman, R. (2019). Resilient food supply chain design: Modelling framework and metaheuristic solution approach. *Computers & Industrial Engineering, 135*, 177–198. https://doi.org/10.1016/j.cie.2019.05.011.

Brettel, M., Friederichsen, N., Keller, M., & Rosenberg, M. (2014). How Virtualization, Decentralization and Network Building Change the Manufacturing Landscape: An Industry 4.0 Perspective. International Journal of Mechanical, Aerospace, Industrial, Mechatronic and Manufacturing Engineering, 8(1), 37–44. Retrieved from http://www.waset.org/publications/9997144.

Cha-Ume, K., & Chiadamrong, N. (2018). Meta-prediction model for introducing lateral transshipment policies in a retail supply chain network through regression analysis. *European Journal of Industrial Engineering, 12*(2), 199–232. https://doi.org/10.1504/EJIE.2018.090615.

Chen, Y., Mockus, L., Orcun, S., & Reklaitis, G. V. (2012). Simulation-optimization approach to clinical trial supply chain management with demand scenario forecast. *Computers and Chemical Engineering, 40*, 82–96. https://doi.org/10.1016/j.compchemeng.2012.01.007.

Cheng, F., Lee, Y. M., Ding, H. W., Wang, W., & Stephens, S. (2008). Simulating Order Fulfillment and Supply Planning for a Vertically Aligned Industry Solution Business. In Proceedings of the 40th Conference on Winter Simulation (pp. 2609–2615). Winter Simulation Conference.

Costa, C, & Santos, M. Y. (2017b). The SusCity big data warehousing approach for Smart Cities. In ACM International Conference Proceeding Series (Vol. Part F1294, pp. 264–273). 10.1145/3105831.3105841.

Costa, E., Costa, C., & Santos, M. Y. (2017). Efficient big data modelling and organization for hadoop hive-based data warehouses. In European, Mediterranean, and Middle Eastern Conference on Information Systems, EMCIS 2017, Lecture Notes in Business Information Processing (Vol. 299, pp. 3–16). 10.1007/978-3-319-65930-5_1.

Costa, Carlos, & Santos, M. Y. (2018). Evaluating Several Design Patterns and Trends in Big Data Warehousing Systems. In 30th International Conference on Advanced Information Systems Engineering, CAiSE 2018, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (Vol. 10816 LNCS, pp. 459–473). 10.1007/978-3-319-91563-0_28.

Costa, C., & Santos, M. Y. (2017a). Big Data: State-of-the-art concepts, techniques, technologies, modeling approaches and research challenges. *IAENG International Journal of Computer Science, 44*(3), 285–301.

Dev, N. K., Shankar, R., Dey, P. K., & Gunasekaran, A. (2014). Holonic supply chain: A study from family-based manufacturing perspective. *Computers & Industrial Engineering, 78*, 1–11. https://doi.org/10.1016/j.cie.2014.09.017.

Duan, C., Deng, C., Gharaei, A., Wu, J., & Wang, B. (2018). Selective maintenance scheduling under stochastic maintenance quality with multiple maintenance actions. *International Journal of Production Research, 56*(23), 7160–7178. https://doi.org/10.1080/00207543.2018.1436789.

Ehmke, J. F., Großhans, D., Mattfeld, D. C., & Smith, L. D. (2011). Interactive analysis of discrete-event logistics systems with support of a data warehouse. *Computers in Industry, 62*(6), 578–586. https://doi.org/10.1016/j.compind.2011.04.007.

Er Kara, M., Oktay Fırat, S.Ü., & Ghadge, A. (2018). A data mining-based framework for supply chain risk management. *Computers & Industrial Engineering*. https://doi.org/10.1016/j.cie.2018.12.017.

Finke, G. R., Schmitt, A. J., & Singh, M. (2010). Modeling and Simulating Supply Chain Schedule Risk. In Proceedings of the Winter Simulation Conference (pp. 3472–3481). Winter Simulation Conference.

Fornasiero, R., Macchion, L., & Vinelli, A. (2015). Supply chain configuration towards customization: A comparison between small and large series production. *IFAC-PapersOnLine, 28*(3), 1428–1433. https://doi.org/10.1016/j.ifacol.2015.06.287.

Gansterer, M. (2015). Aggregate planning and forecasting in make-to-order production systems. *International Journal of Production Economics, 170*, 521–528. https://doi.org/10.1016/j.ijpe.2015.06.001.

Ghadge, A., Dani, S., & Kalawsky, R. (2012). Supply chain risk management: Present and future scope. *The International Journal of Logistics Management, 23*(3), 313–339. https://doi.org/10.1108/09574091211289200.

Gharaei, A., Hoseini Shekarabi, S. A., & Karimi, M. (2019). Modelling And optimal lot-sizing of the replenishments in constrained, multi-product and bi-objective EPQ models with defective products: Generalised Cross Decomposition. *International Journal of Systems Science: Operations & Logistics, 1–13*. https://doi.org/10.1080/23302674.2019.1574364.

Gharaei, A., Karimi, M., & Hoseini Shekarabi, S. A. (2019b). Joint economic lot-sizing in multi-product multi-level integrated supply chains: Generalized benders decomposition. *International Journal of Systems Science: Operations & Logistics, 1–17*. https://doi.org/10.1080/23302674.2019.1585595.

Gharaei, A., Karimi, M., & Hoseini Shekarabi, S. A. (2019a). An integrated multi-product, multi-buyer supply chain under penalty, green, and quality control polices and a vendor managed inventory with consignment stock agreement: The outer approximation with equality relaxation and augmented penalty algorithm. *Applied Mathematical Modelling, 69*, 223–254. https://doi.org/10.1016/j.apm.2018.11.035.

Giri, B. C., & Bardhan, S. (2014). Coordinating a supply chain with backup supplier through buyback contract under supply disruption and uncertain demand. *International Journal of Systems Science: Operations & Logistics, 1*(4), 193–204. https://doi.org/10.1080/23302674.2014.951714.

Giri, B. C., & Masanta, M. (2018). Developing a closed-loop supply chain model with price and quality dependent demand and learning in production in a stochastic environment. *International Journal of Systems Science: Operations & Logistics, 1–17*. https://doi.org/10.1080/23302674.2018.1542042.

Goss, R. G., & Veeramuthu, K. (2013). Heading towards big data building a better data warehouse for more data, more speed, and more users. In ASMC 2013 SEMI Advanced Semiconductor Manufacturing Conference (pp. 220–225). IEEE. 10.1109/ASMC.2013.6552808.

Grover, P., & Kar, A. K. (2017). Big data analytics: A review on theoretical contributions and tools used in literature. *Global Journal of Flexible Systems Management, 18*(3), 203–229. https://doi.org/10.1007/s40171-017-0159-3.

Ho, W., Zheng, T., Yildiz, H., & Talluri, S. (2015). Supply chain risk management: A literature review. *International Journal of Production Research, 53*(16), 5031–5069. https://doi.org/10.1080/00207543.2015.1030467.

Hoseini Shekarabi, S. A., Gharaei, A., & Karimi, M. (2018). Modelling and optimal lot-sizing of integrated multi-level multi-wholesaler supply chains under the shortage and limited warehouse space: Generalised outer approximation. *International Journal of Systems Science: Operations & Logistics, 1–21*. https://doi.org/10.1080/23302674.2018.1435835.

Jahangirian, M., Eldabi, T., Naseer, A., Stergioulas, L. K., & Young, T. (2010). Simulation in manufacturing and business: A review. *European Journal of Operational Research, 203*(1), 1–13. https://doi.org/10.1016/j.ejor.2009.06.004.

Kagermann, H., Helbig, J., Hellinger, A., & Wahlster, W. (2013). Recommendations for Implementing the Strategic Initiative INDUSTRIE 4.0: Securing the Future of German Manufacturing Industry ; Final Report of the Industrie 4.0 Working Group. Forschungsunion. Retrieved from < https://books.google.pt/books?id=AsfOoAEACAAJ > .

Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013). Big Data: Issues and Challenges Moving Forward. In 2013 46th Hawaii International Conference on System Sciences (pp. 995–1004). IEEE. 10.1109/HICSS.2013.645.

Kırılmaz, O., & Erol, S. (2017). A proactive approach to supply chain risk management: Shifting orders among suppliers to mitigate the supply side risks. *Journal of Purchasing and Supply Management, 23*(1), 54–65. https://doi.org/10.1016/j.pursup.2016.04.002.

Kv, R. S., & Kavya, N. P. (2016). Trend analysis of e-commerce data using Hadoop ecosystem. *International Journal of Computer Applications, 147*(6), 1–5.

Lasi, H., Fettke, P., Kemper, H.-G., Feld, T., & Hoffmann, M. (2014). Industry 4.0. *Business and Information Systems Engineering, 6*(4), 239–242. https://doi.org/10.1007/s12599-014-0334-4.

Lee, Y. M., Ghosh, S., & Ettl, M. (2009). Simulating distribution of emergency relief supplies for disaster response operations. In Winter Simulation Conference (pp. 2797–2808). Winter Simulation Conference.

Longo, F., & Mirabelli, G. (2008). An advanced supply chain management tool based on modeling and simulation. *Computers and Industrial Engineering, 54*(3), 570–588. https://doi.org/10.1016/j.cie.2007.09.008.

Lourenço, J. R., Abramova, V., Vieira, M., Cabral, B., & Bernardino, J. (2015). NoSQL Databases: A Software Engineering Perspective (pp. 741–750). 10.1007/978-3-319-16486-1_73.

Macchion, L., Fornasiero, R., & Vinelli, A. (2017). Supply chain configurations: A model to evaluate performance in customised productions. *International Journal of Production Research, 55*(5), 1386–1399. https://doi.org/10.1080/00207543.2016.1221161.

Madden, S. (2012). From databases to big data. *IEEE Internet Computing, 16*(3), 4–6. https://doi.org/10.1109/MIC.2012.50.

Masoud, S. A., & Mason, S. J. (2016). Integrated cost optimization in a two-stage, automotive supply chain. *Computers and Operations Research, 67*, 1–11. https://doi.org/10.1016/j.cor.2015.08.012.

Matsuo, H. (2015). Implications of the Tohoku earthquake for Toyota's coordination mechanism: Supply chain disruption of automotive semiconductors. *International Journal of Production Economics, 161*, 217–227. https://doi.org/10.1016/j.ijpe.2014.07.010.

Mishra, M., & Chan, F. T. S. (2012). Impact evaluation of supply chain initiatives: A system simulation methodology. *International Journal of Production Research, 50*(6), 1554–1567. https://doi.org/10.1080/00207543.2011.556151.

Mohanty, S., Jagadeesh, M., & Srivatsa, H. (2013). Big data imperatives: Enterprise 'Big Data'warehouse,'BI'implementations and analytics. Apress.

Nageshwaraniyer, S. S., Meng, C., Maghsoudi, A., Son, Y.-J., & Dessureault, S. (2012). Simulation-based decision support system for sustainable coalmining operations. In 62nd IIE Annual Conference and Expo 2012 (pp. 1574–1583).

Nodarakis, N., Sioutas, S., Tsakalidis, A., & Tzimas, G. (2016). Using Hadoop for Large Scale Analysis on Twitter: A Technical Report. ArXiv Preprint ArXiv:1602.01248. Retrieved from http://arxiv.org/abs/1602.01248.

Park, Y., Hong, P., & Roh, J. J. (2013). Supply chain lessons from the catastrophic natural disaster in Japan. *Business Horizons, 56*(1), 75–85. https://doi.org/10.1016/j.bushor.2012.09.008.

Pires, B., Goldstein, J., Higdon, D., Korkmaz, G., Keller, S., Shipp, S., ... Koehler, A. (2016). *A Bayesian Simulation Approach for Supply Chain Synchronization* (pp. 3698–3699). Piscataway, NJ, USA: IEEE Press.

Ponte, B., Sierra, E., de la Fuente, D., & Lozano, J. (2017). Exploring the interaction of inventory policies across the supply chain: An agent-based approach. *Computers & Operations Research, 78*, 335–348. https://doi.org/10.1016/j.cor.2016.09.020.

Posada, J., Toro, C., Barandiaran, I., Oyarzun, D., Stricker, D., De Amicis, R., ... Vallarino, I., Jr. (2015). Visual computing as a key enabling technology for Industrie 4.0 and industrial internet. *IEEE Computer Graphics and Applications, 35*(2), 26–40. https://doi.org/10.1109/MCG.2015.45.

Rabe, M., & Dross, F. (2015). A reinforcement learning approach for a decision support system for logistics networks. In 2015 Winter Simulation Conference (WSC) (pp. 2020–2032). IEEE. 10.1109/WSC.2015.7408317.

Robinson, S. (2005). Discrete-event simulation: From the pioneers to the present, what next? *Journal of the Operational Research Society, 56*(6), 619–629. https://doi.org/10.1057/palgrave.jors.2601864.

Rocha de Paula, M., Boland, N., Ernst, A. T., Mendes, A., & Savelsbergh, M. (2019). Throughput optimisation in a coal export system with multiple terminals and shared resources. *Computers & Industrial Engineering, 134*, 37–51. https://doi.org/10.1016/j.cie.2019.05.021.

Sahoo, K., & Mani, S. (2016). GIS based discrete event modeling and simulation of biomass supply chain. In Proceedings – Winter Simulation Conference (pp. 967–978). 10.1109/WSC.2015.7408225.

Santos, M. Y., Oliveira e Sá, J., Andrade, C., Vale Lima, F., Costa, E., Costa, C., ... Galvão, J. (2017). A Big Data system supporting Bosch Braga Industry 4.0 strategy. International Journal of Information Management, 37(6), 750–760. 10.1016/j.ijinfomgt.2017.07.012.

Sarkar, S., & Giri, B. C. (2018). Stochastic supply chain model with imperfect production and controllable defective rate. *International Journal of Systems Science: Operations & Logistics, 1–14*. https://doi.org/10.1080/23302674.2018.1536231.

Schmitt, A. J., & Singh, M. (2009). Quantifying supply chain disruption risk using monte carlo and discrete-event simulation. In Winter Simulation Conference (pp. 1237–1248). Winter Simulation Conference.

Schwede, C., Sieben, B., Song, Y., Hellingrath, B., & Wagenitz, A. (2009). A simulation-based method for the design of supply strategies to enter developing markets. *International Journal of Simulation and Process Modelling, 5*(4), 324–336. https://doi.org/10.1504/IJSPM.2009.032595.

Shah, N. H., Chaudhari, U., & Cárdenas-Barrón, L. E. (2018). Integrating credit and replenishment policies for deteriorating items under quadratic demand in a three echelon supply chain. *International Journal of Systems Science: Operations & Logistics, 1–12*. https://doi.org/10.1080/23302674.2018.1487606.

Simchi-Levi, David, Kaminsky, P., Simchi-Levi, E., & Shankar, R. (2008). *Designing and managing the supply chain: Concepts, strategies and case studies.* Tata: McGraw-Hill Education.

Simchi-Levi, D., Schmidt, W., Wei, Y., Zhang, P. Y., Combs, K., Ge, Y., ... Zhang, D. (2015). Identifying risks and mitigating disruptions in the automotive supply chain. *Interfaces, 45*(5), 375–390. https://doi.org/10.1287/inte.2015.0804.

Thun, J.-H., & Hoenig, D. (2011). An empirical analysis of supply chain risk management in the German automotive industry. *International Journal of Production Economics, 131*(1), 242–249. https://doi.org/10.1016/j.ijpe.2009.10.010.

Thusoo, A., Sarma, J. Sen, Jain, N., Shao, Z., Chakka, P., Zhang, N., … Murthy, R. (2010). Hive – a petabyte scale data warehouse using Hadoop. In 2010 IEEE 26th International Conference on Data Engineering (ICDE 2010) (pp. 996–1005). IEEE. 10.1109/ICDE.2010.5447738.

Thusoo, A., Shao, Z., Anthony, S., Borthakur, D., Jain, N., Sen Sarma, J., … Liu, H. (2010). Data warehousing and analytics infrastructure at facebook. In Proceedings of the 2010 international conference on Management of data - SIGMOD '10 (p. 1013). New York, New York, USA: ACM Press. 10.1145/1807167.1807278.

Tiwari, S., Wee, H. M., & Daryanto, Y. (2018). Big data analytics in supply chain management between 2010 and 2016: Insights to industries. *Computers & Industrial Engineering, 115*, 319–330. https://doi.org/10.1016/j.cie.2017.11.017.

Turner, C. J., Hutabarat, W., Oyekan, J., & Tiwari, A. (2016). Discrete event simulation and virtual reality use in industry: New opportunities and future trends. *IEEE Transactions on Human-Machine Systems, 46*(6), 882–894. https://doi.org/10.1109/THMS.2016.2596099.

Ularu, E. G., Puican, F. C., Apostu, A., & Velicanu, M. (2012). Perspectives on big data and big data analytics. *Database Systems Journal, 3*(4), 3–14.

Vieira, A. A., Dias, L. M., Santos, M. Y., Pereira, G. A., & Oliveira, J. A. (2018). Setting an industry 4.0 research and development agenda for simulation – A literature review. *International Journal of Simulation Modelling, 17*(3), 377–390. https://doi.org/10.2507/IJSIMM17(3)429.

Vieira, António A. C., Pedro, L., Santos, M. Y., Fernandes, J. M., & Dias, L. S. (2018). Data Requirements Elicitation in Big Data Warehousing. In European, Mediterranean, and Middle Eastern Conference on Information Systems, EMCIS 2018, Lecture Notes in Business Information Processing (Vol. 341, pp. 106–113). 10.1007/978-3-030-11395-7_10.

Vieira, A. A. C., Dias, L. M. S., Santos, M. Y., Pereira, G. A. B., & Oliveira, J. A. (2018). Simulation of an Automotive Supply Chain in Simio: Data Model Validation. In 30th European Modeling and Simulation Symposium, EMSS (pp. 294–301).

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. https://doi.org/10.1080/07421222.1996.11518099.

Yin, S., Nishi, T., & Zhang, G. (2016). A game theoretic model for coordination of single manufacturer and multiple suppliers with quality variations under uncertain demands. *International Journal of Systems Science: Operations & Logistics, 3*(2), 79–91. https://doi.org/10.1080/23302674.2015.1050079.

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 and Industrial Engineering, 101*, 572–591. https://doi.org/10.1016/j.cie.2016.07.013.

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