ISSN: 2366-1313

Case Study and Systematic Analysis of Big Data Analytics for E-Commerce Logistics

G Radhika, Research Scholar, Department of Computer Science , J.S University, Shikohabad.

Dr. Badarla Anil ,Professor ,Supervisor, Department of Computer Science, J.S University, Shikohabad.

Abstract - Because of the ways in which the digital revolution has altered the expectations of consumers and the ways in which they make purchases, e-commerce has seen enormous growth. This development brought to an environment that was active and chaotic, with a rising density of distribution networks. This ecosystem consists of a significant number of distribution sites, a variety of delivery methods, and various expectations on last-mile delivery. It is possible that the implementation of big data analytics would prove to be very profitable in the context of e-commerce supply chain management. This is due to the complexity and proliferation of data issues. In spite of the increased interest, there are still relatively few research that investigate how big data may be used. We defined the current practices of e-commerce logistics using a design science methodology, as well as those that are projected and may, to a considerable part, be supported by proper big data technology. This was done so that we could compare and contrast the two sets of practices. We came up with a list of business criteria that describes the need for e-commerce logistics, and we used those criteria to arrive at our conclusion. After that, the requirements were reimagined in the form of a number of use case scenarios, which were designed to illustrate the potential benefits of using big data analytics. As a last suggestion, we will discuss a potential design for a big data analytics artifact that could be able to meet the requirements of e-commerce supply chain management.

Keywords: big data, business analytics, e-commerce logistics, collaboration, requirements

1. INTRODUCTION

Customers' expectations have been altered, and as a result, their shopping habits have changed. At the same time, the number of people who use the internet and who buy things online continues to rise. 57% of European Internet users presently make purchases online, according to research conducted by E-commerce Europe (2016). This indicates that e-commerce has been responsible for practically the entire expansion of the retail sector over the last 10 years. For instance, since 2011, sales in European business-to-consumer e-commerce have consistently climbed, culminating in a European business-to-consumer e-commerce turnover of €509.9 billion (E-commerce Europe, 2016). Customers have a growing expectation that shops would provide a high level of service regardless of the method through which they make purchases or get

ISSN: 2366-1313

deliveries of goods. According to research conducted by E-commerce Europe (2016), forty percent of consumers want it to be even easier to make purchases online and offline, forty one percent want improved customer service, forty five percent want faster delivery, and forty six percent want simplified return and refund processes. To increase sales across all channels and platforms, as well as to keep their promises about product availability and delivery, retailers need to ensure that their customers have a consistent and positive experience across all of their touchpoints.

During this time period, merchants are facing additional issues relating to inventory visibility, channel integration, delivery speed, and the ease of returning items purchased. Every retailer, regardless of whether they do business only online, in physical locations, or both, is subject to the same kind of difficulties. According to research conducted in the United Kingdom and Germany (E-commerce Europe, 2016), more than 65 percent of clients who are under the age of 21 want to have both a physical and an online experience. The growing variety of channel formats and the move from single to multi and ultimately omni-channel formats have resulted in enhanced browsing and buying experiences for customers. However, logisticians, including both upstream suppliers and downstream merchants, are finding it more challenging to manage these changes. There has also been an increase in the formation of partnerships in which conventional stores with physical locations pool their resources and networks with those of online retailers. Amazon just completed the acquisition of Whole Foods, bringing its presence farther into the real world (Aziza, 2017). This new trend will make the existing structure of the market more unstable, which will make it necessary to improve both the effectiveness and efficiency of logistics. According to E-commerce Europe (2016), retail strategies that successfully handle this complexity will have more success throughout the course of their long-term implementation.

The emergence of e-commerce has ushered in a new age, and with it comes a plethora of new developments, one of which is the significant disruption now being experienced in the transportation sector. According to Boyer (2009) and Hubner et al. (2016), the density of the distribution network has a significant impact on the cost of distribution. The distribution network density comprises the number of delivery locations, the number of delivery routes, and the requirements for last-mile delivery. This is due to the fact that e-commerce and, more recently, the rise of omni-channels have produced a market that is both dynamic and turbulent. Customers have begun to expect more dynamic last-mile delivery services, which has resulted in stores being turned into fulfilment hubs where customers may pick up online purchases, thereby making the process more complicated (Jeseke, 2013). Home delivery services are quite popular among online shoppers; nevertheless, this fragments the flow of packages, which makes things more difficult in metropolitan areas. The fact that there are several distribution needs, such as orders and returns from a range of customers at different stages, has also contributed to the complexity of this situation. As a direct result of this, e-commerce logistics has emerged as a crucial area for innovation and is now one of the key policy concerns being addressed by the European Commission (E-commerce Europe, 2016).

.

2. RESEARCHBACKGROUND

According to Ecommerce Europe (2016), the expansion of e-commerce results in the emergence of new business processes and patterns of customer behavior. As a result of the advantages and drawbacks brought

ISSN: 2366-1313

about by the information revolution, big data analytics, also known as BDA, have become an increasingly important component of online commerce during the last several years. Big data analytics, also known as BDA, is a technique that may be used in e-commerce to enhance decision-making across all processes, such as infrastructure and operations on the one hand, and customer behavior and happiness on the other (Miller, 2013). This creates the opportunity for a more optimal equilibrium between supply and demand. The term "big data" refers to the process of gathering vast amounts of data from a variety of sources, as well as processing and analyzing that data in real time. Akter et al. (2016) provides a comprehensive assessment of big data analysis as it relates to e-commerce. According to Akter et al. (2016), Koutsabasis et al. (2008), and Mehra (2013), some of the most significant application areas for big data in the e-commerce industry include personalized services, dynamic pricing, predictive analytics, supply chain optimization, and visibility. According to McAfee and Brynjolfsson (2012), businesses that use BDA in their operations claim an improvement in productivity that ranges from 5 to 6 percent.

According to Swaminathan (2012), e-commerce logistics applications and trends might potentially benefit from big data analytics. According to Jeseke et al. (2013), the collection and analysis of large amounts of data may be of assistance to decision-makers in three primary areas of logistical application: the customer experience; operational efficiency and network design; and new business models. The significance of businesses recognizing Big Data Analytics as a strategic asset that has to be understood and integrated holistically has been highlighted in prior research (Addo-Tenkorang and Helo, 2016), as has the function that it plays in the supply chain (Wang et al., 2016). There are certain use cases that have been discovered in the research that make use of the usefulness of big data analytics in relation to the distribution process in order to improve operational performance. Some of the most notable uses of big data in the logistics industry include real-time route optimization (Pillac et al., 2013; Fabian and Christian, 2012), demand forecasting, crowd-based pick-up and delivery (Jeseke et al., 2013), and dynamic inventory routing (Sage, 2013; Wang et al., 2016). Other applications include crowd-based pick-up and delivery (Jeseke et al., 2013). According to the conclusions of the evaluation of the relevant literature, there is the potential to improve logistical operations, particularly those pertaining to e-commerce logistics, by making use of BDA.

3. RESEARCH APPROACH

In accordance with the aforementioned aims and objectives of the study, we have chosen to construct our methodology on the basis of the design science paradigm (Jones and Gregor, 2007; Hevner et al., 2004; Peffers, 2007). According to Berente and Lyytinen (2006), this methodology makes an effort to resolve issues by using novel and creative methods to produce artifacts, which then put into practice and evaluate various processes and ideas throughout the iterative development process.

We started off by providing an overview of the particular study issue and argued that big data analytics is beneficial for e-commerce supply chain management. A literature study and semi-structured interviews with experts in the field of business practices were used to identify the gaps in the existing research. From the 20th of February through the 10th of March in 2017, a series of in-depth interviews were carried out at five different companies in order to have a

ISSN: 2366-1313

better understanding of the industry's expectations. The companies that agreed to be interviewed were chosen according to two criteria: the nature of their company (online-only vs click-and-mortar operations), and whether or not they collaborated with a third-party logistics provider. In order to conduct the interviews, a set of guidelines known as a "interview protocol" were devised. Research was conducted on the following subjects using the protocol: a) the demographics of the companies; b) an overview of the current state of e-commerce logistics; c) potential future scenarios for e-commerce logistics and the role of big data technologies; and d) attitudes towards collaborative logistics practices in the e-commerce sector. Among those that responded were the leading courier service in Greece, a third-party logistics (3PL) firm that specializes in e-commerce logistics, and three merchants of electronic goods who have a presence in every nation (one who sells only online and two who sell both online and in physical stores). This technique revealed a number of problematic areas in existing practices, as well as some drivers and obstacles to the adoption of collaborative logistics practices, as well as numerous issues that should be addressed for both the design and implementation of big data analytics artifacts. In addition, this method highlighted a number of problematic areas in current practices, as well as some drivers and hurdles to the adoption of collaborative logistics practices. The repercussions of reaching this level are discussed in Section 4.1.

Through analyzing and compiling the information gleaned from the method that came before it, we were able to come at a list of high-level business needs, which are discussed in Section 4.2. In order to get a deeper comprehension of the business goals and proceed with the development of the big data analytics artifact (Section 4.3), we devised a collection of use case scenarios that fulfill the aforementioned business goals. These scenarios were designed to meet the needs of the company. The research's industrial setting, which included manufacturers, suppliers, traditional brick and mortar stores, and internet merchants, served as the inspiration for the creation of the scenarios. We worked in close collaboration with two representative retailers, an offline store and an online store, as well as a firm that specializes in third-party logistics (3PL), in order to determine the particular goals of the big data analytics artifact and proceed with the design and validation of the proposed solution. The physical shop is part of a large retail chain in Greece that has 240 locations (190 supermarkets and 50 CashandCarry stores), as well as a supply network that is formed of three main warehouses. One of the first and largest online merchants in Greece, and also the largest and most successful retailer in Greece when it comes to e-commerce logistics. After that, we determine the goals of a big data analytics artifact by utilizing the issue statement in conjunction with our understanding of what is possible and what can be accomplished.

The creation of the item as well as its design followed next. The artifact of the research is an analytics infrastructure for large amounts of data. The method of designing and developing the system is exactly the same as that of an information system. As was noted earlier, it began with a technique for collecting requirements, in which a large number of possible users took part. This produced requirements documentation, which was then used in the process of designing a conceptual architecture (details of which are provided in Section 5).

4. BUSINESS REQUIREMENTS AND USE CASE SCENARIOS IN THE E-COMMERCE LOGISTICS

ISSN: 2366-1313

- **In-depth interviews and literature review in sights**

In this part of the report, we will provide the information that we gained from the in-depth interviews and explore how it ties to the conclusions that were found in the literature study. It is vital to stress that the data obtained during the interviews and the findings of the literature coincided. This is because the firms chosen represent not only local but also international or extroverted enterprises. As a consequence of this, despite the difficulties that are specific to each nation, we may reach the conclusion that the conditions for doing business that are mentioned in the next section are sufficiently broad. To begin, respondents emphasized that big data technologies have the potential to gather information from a variety of sources (including social media, sales, location, and geolocation data) and produce solutions that can lower delivery costs while simultaneously improving the efficacy of distribution processes. Integration of data sources, analytical approaches, and business experience is required in order for big data to be used effectively in the logistics industry. According to the findings of the study, analytics based on big data have a significant potential to enhance operational effectiveness. Both the design of the network and the optimization of the final mile are components of operational efficiency. It is possible that real-time optimization and integrated pick-up and delivery will considerably increase the efficiency of the final mile. Network planning is a second field of big data applications in logistics that is connected to operational efficiency (Jeseke et al., 2013). Network planning comprises choices on warehouses, distribution centers, and vehicles that are created specifically for the company.

Throughout the course of the interviews, the most important features and problematic aspects of the business of e-commerce and logistics were emphasized. During the first interviews, it became clear that the e-commerce industry had a high level of competition, which is one of the most important factors preventing the widespread adoption of collaborative logistics strategies and the sharing of data among supply chain partners. The primary problem areas that have been identified are the nation's geographic diversity and low population density, the problematic zip code system, the national orientation of the majority of retailers' companies, certain unique characteristics of the majority of 3PL companies, the effects of the ongoing fiscal and economic crisis, and a number of legislative constraints. In addition, there are a number of potential solutions that have been proposed. The high costs of distribution are a direct result of the aforementioned elements, and the costs continue to escalate until the last mile of the delivery process.

All of the respondents agreed that the last mile delivery is the most difficult and costly part of the delivery process. This is due to the fact that it has a direct impact on how customers evaluate a company. This cost is enhanced in the Greek market due to the phenomena of "cash on delivery," which is selected by more than 75% of online clients. This choice results in an elevated cost. If the customer chooses this mode of payment, they are able to use their right to refuse delivery; as a consequence, there will be a greater number of items that have not been delivered and will need to be sent back. When analyzing the factors that contribute to the high costs of last-mile delivery, one further factor to take into account is the oligopoly of courier services that exists in the Greek market. The remaining forty percent of the market share is held by one of the major businesses, and it is distributed among the other five delivery companies. Couriers, on the other hand, play an essential part in the supply chain of e-commerce logistics since they are responsible for the last leg of distribution as well as the expenses associated with that leg. In addition to this, they may provide a number of dynamic delivery choices for customers to choose from, as well as order traceability data that enables a wide range of sophisticated analytics for use in e-commerce supply chain management.

According to Pillac et al. (2013) and Fabian and Christian (2012), using real-time data from several sources (such as traffic data, sensors, and real-time events) to dynamically optimize routing and provide drivers with instant instructions may significantly enhance last mile distribution. According to Du et al.'s (2005) research, the unpredictable and ever-changing nature of an online store's orders makes dynamic routing a much more difficult task. With dynamic vehicle routing, it is possible to take into consideration demands for both delivery and fresh return. According to the opinions of three out of every five respondents, re-distribution or returns systems are among the e-commerce logistics activities that are the most difficult and expensive to carry out. As a result of this, dynamic rerouting that is based on the merging of deliveries and fresh return requests may increase the operational efficiency of the business. In addition, if you're using a Vendor Managed Inventory (VMI) system, dynamic routing might be beneficial for inventory delivery depending on the sales data you've been collecting. It is possible to apply dynamic inventory routing in this scenario since the routing is determined on a more temporary timescale. Special purpose models may be of great assistance when it comes to streamlining the routing plan, refining future demand projections, and reducing the need for new (non-prescheduled) vehicle routes and the costs associated with them (Wang et al., 2016; Sage 2013). These are all goals that may be accomplished by using the models.

The "reverse logistics" technique is an essential part of the process of logistics for online commerce since European legislation permits buyers to return products within 14 days after the purchase date. A different proportion of total return applies to each market sector. It should not come as a surprise that food and specialty retailers have a low proportion of repeat customers. Because customers already know what they want when they shop at a specialized store, it may be difficult for them to return fresh things like food that has been opened. This is likewise the case for the businesses in the electronic sector that were represented by the individuals we interviewed. When compared to many other fields, the fashion industry has a very high return rate. The vast selection of available sizes, materials, and other attributes in clothing is the fundamental driver behind this phenomenon. Consider the garment business as an example; it has a total return rate of 16.50%. On the other hand, there are businesses that are of the opinion that a high return rate is not always a negative thing as long as it is tied to the main purpose of the firm. Nevertheless, their collecting makes distribution techniques more difficult and adds to the expenses of distribution. Many third-party logistics (3PL) organizations struggle with fundamental challenges such as reverse logistics and empty runs. Even if a truck has a high and desirable load factor, returning empty at the end of the day will still result in expenses (regardless of whether the companies themselves pay them or not). As a direct result of this, the use of big data technology in the field of returns was considered a workable alternative.

- **Business requirements**

In accordance with the methodology detailed in the part that came before this one, we have produced a list of business standards that outline the requirements for

ISSN: 2366-1313

logistics pertaining to online commerce. Following is a list of the major criteria that were selected after taking into consideration factors such as perceived value and possible advantages. Our mission is to study the ways in which big data analytics solutions may be able to meet these business goals, to suggest a big data analytics architecture, and to locate any possible barriers.

- **Req. 1: To explore how collaboration could be applied in the e-commerce logistics domain in order to address the current challenges and support firms to decrease cost and improve the overall distribution process performance.**

It would seem that cooperation amongst participants in the supply chain is one of the aspects of e-commerce logistics that presents the greatest challenge. The difficulties associated with last-mile delivery, the demand for a supply chain that is more responsive, and the requirement for cost savings have all contributed to the prominence of collaborative solutions. According to Tyan et al. (2003) and Danloup et al. (2014), unified pick-up and delivery operations that combine the paths of many supply chain stakeholders may boost operational efficiency in some scenarios, such as those involving the final mile of a supply chain. The interviews indicate that e-commerce businesses are reluctant to share data and common vehicles, despite the fact that doing so might be the key to reducing delivery costs. They need further proof that demonstrates the worth of the cooperation.

- **Req. 2: To identify potential synergies among the e-commerce stakeholders at the reverse logistics.**

According to both academic research and interviews conducted inside the business, reverse logistics is one of the processes that takes up the most time and results in the highest costs associated with e-commerce logistics. Costs and the usual return rate may be quite different from one business to the next (for example, the garment industry and the electronics industry), and the return rate in the fashion industry can reach as high as thirty percent.

- **Req. 3: To analyse current distribution processes in order to depict the current distribution patterns and forecast future problems.**

The abundance of data (routing data, order data, and customer data), when combined with an approach that analyzes big data, can assist in the identification of distribution patterns, the improvement of distribution process visualization, and the support of the identification of problematic processes, all of which will contribute to an increase in operational efficiency.

- **Req. 4: To provide alternative shipping methods at the consumer in order to increase customers' satisfaction and provide lower prices.**

The use of alternative last-mile delivery strategies, such as click and collect and pick up locations, has the potential to save costs for retailers, third-party logistics providers (3PLs), and courier services while simultaneously improving customer service and the overall experience for customers. Finding locations for pick-up and drop-off is one of the most important requirements for last-mile optimization.

- **Req. 5: To enable dynamically changing supply chains that take into consideration various routing and customer preference characteristics in order to decrease logistics costs and to increase customers' satisfaction.**

According to the interviews, the courier services have identified the need of consumers to dynamically adjust the time of delivery as well as the location of the delivery, and they are already implementing new services in this area. Real-time dynamic route optimization is one of the most important requirements for e-Commerce logistics. This kind of optimization makes use of real-time data from a variety of sources (such as traffic statistics, sensors, and real-time events) to dynamically enhance routing (Pillac et al., 2013; Fabian and Christian, 2012).

- **Req. 6: To identify users' problems relevant with the logistics procedure, via the analysis of social media data.**

The expansion of social media platforms presents the opportunity to make use of the open data that these platforms make available. By leaving comments on blogs, forums, and social networking sites, Internet users communicate their opinions, regardless of whether those opinions are good or negative. It is possible to do an analysis on these data in order to acquire knowledge and determine challenges associated with the logistics process. It is possible to analyze the data and obtain important information by using algorithms that are based on meaning approaches and text mining.

➢ **Use case scenarios**

We construct a set of use case scenarios that takes into consideration both the needs of business and the value that is recognized in order to better understand the role that big data analytics plays in tackling the issues that are presented by e-commerce logistics. Following is a quick discussion of each of the use cases, along with the potential contribution that big data analytics might make to their respective implementations. Use case options are shown in Figure 1, along with how these possibilities link to business needs.

- **Use Case Scenario 1: End to end information sharing in the supply chain: An integrated view of supply chain towards collaboration of a 3PL and a courier company.**

The primary objective of this scenario is to map the whole of the logistics process by making use of the data provided by couriers and 3PL companies (including delivery, product, fleet, and route information) in order to look for ways in which distribution operations may be made more efficient. It is possible that rerouting may increase the operational efficiency of distribution operations. This will be accomplished by dynamically integrating fresh return requests and delivery requests. Big data technology may gather information in real time from a range of sources (including 3PL businesses, couriers, and open data), and big data analytics, in the form of routing algorithms, can give solutions that can lower delivery costs while simultaneously boosting the efficacy of distribution processes.

- **Use Case Scenario 2: Shared micro-hubs among online retailers.**

Shared micro-hubs provide an alternative to traditional distribution strategies for businesses operating in the e-commerce space. It's possible that the shared warehouses will work as a center for a number of different stores, logistics providers, and online businesses, acting as the first point of contact for the delivery of last-mile products to customers. The combined efforts provide a logistical network that is both flexible and efficient for all parties involved, and they have the potential to significantly improve the logistics of (urban) e-commerce by reducing the amount of time spent traveling, the amount of money spent on distribution, and the amount of carbon

ISSN: 2366-1313

dioxide emissions produced. In this scenario, the information from a number of different online retailers and 3PL companies may be utilised. The establishment of micro-hub sites is essential to the enhancement of distribution operations.

- **Use Case Scenario 3: Shared reverse logistics among online retailers.**

This use case scenario illustrates how internet merchants may work together to improve their business. This scenario proposes to lower the cost of return collection as well as CO_2 emissions by combining distribution flows from client delivery locations and deploying a single truck to collect returns through common or near delivery sites. Additionally, this scenario will combine distribution flows from client delivery locations, which will result in less CO_2 emissions. This scenario will also make use of the concept of dynamic re-routing in order to show other locations at which the product may be picked up when a courier is in the area in order to distribute products to various locations.

- **Use Case Scenario 4: Identification of delivery patterns and problematic issues and future forecasting.**

The fourth possible scenario involves determining the geographical and temporal distribution patterns of certain e-commerce merchants in order to improve inventory management and make the delivery procedure more straightforward. It is feasible to identify behavior patterns and particular issues with distribution operations (missing deliveries) by analyzing historical data, seasonal trends, and new freight movements. This may be done by comparing the three types of data. In addition, precise demand forecasting is included in this scenario. Forecasting deliveries by region or season may be important information since it may assist in the optimization of logistical operations in network architecture, such as the placement of distribution centers and warehouses, as well as operational decisions, such as fleet capacity planning. Additionally, this information may help in the event that there is a need to plan for additional fleet capacity.

- **Use Case Scenario 5: Shared Click & Collect Points (in terms of space) where customers collect their online orders.**

Customers that make use of click and collect points buy items online and then go to a collection facility to pick them up after making their purchases. These days, when a shop offers the Click & Collect service, they utilize their own stores as the "collection stations," which means that a journey is still required, but the customer saves time choosing out their items. In this use case, businesses would put up "click and collect" facilities in the town centers of their respective locations, such as post offices or bus or train stations, so that consumers could pick up their purchases there. The concept of shared click-and-collect terminals proposes that these locations might be specialized spaces inside already established shopping malls or smaller warehouses with lockers that are mostly supplied with little packets of merchandise. Off-peak hours are a good time for retailers and couriers to transport customer orders to these locations, where consumers may then pick them up.

- **Use Case Scenario 6: Inventory routing in an omni-channel environment.**

This example of a use case scenario focuses on a potential partnership between a traditional grocery store chain and an online supermarket. The primary objective

ISSN: 2366-1313

of using the supermarket chain's physical network as intermediary hubs, which will be in charge of satisfying local demands, is to cut down on the expenses associated with transportation. The use case will cover the challenges of selecting the most cost-effective physical locations to convert into hubs, as well as the challenges associated with the distribution of commodities between these intermediate nodes and the main hub. Choices need to be taken about the quantities, timing, and routes of each of these hubs in order to bring down the overall cost of the process. Data on routes, fleets, deliveries, and orders are all included in this scenario. The issue that was discussed before is referred to as an Inventory Routing Problem, or IRP for short.

- **Use Case Scenario 7: Online consumer insights regarding logistics and delivery processes.**

The primary objective of this hypothetical situation is to collect information from social media in order to learn about the issues and preferences of online customers that are pertinent to the logistical process. The reviews written by customers fall under the category of "unstructured data" since these evaluations are written in free-form. It is possible to analyze the data and obtain important information by using algorithms that are based on meaning approaches and text mining.

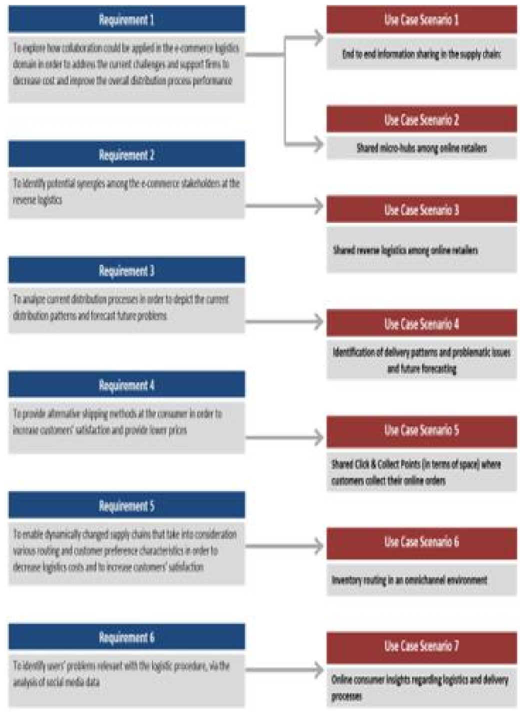

Figure 1- Business requirements and use case scenarios

These specific use case scenarios will be used in order to design the big data analytics artefact.

5. PROPOSED BIG DATA ANALYTICS ARTEFACT DESIGN

We picked a number of various technologies and procedures that would be employed to facilitate the execution of the use case scenarios based on the business needs and use case scenarios that were discussed before. In addition, we will demonstrate how the different technologies and algorithms will be combined into the various layers of a larger architecture (Figure 2), which will be shown in the following paragraphs.

- **Data sources:**

The data sources layer includes the external data sources that will be used in order to retrieve datafor implementing the various use case scenarios. The data

ISSN: 2366-1313

sources include: online retailers' data, 3PLs data,couriers'data andsocialmedia data.

- **Connectors:**

The Core IDS Connector is this layer's most important component; it was developed to facilitate the exchange of data across many manufacturing companies. The Connectors provide an increased level of safety by integrating a number of different security mechanisms, and data will only be sent between parties who have been verified and connected. Data is sent back and forth between the company-side connection and the framework-side connector, both of which are distinct components of the connector's overall infrastructure design. This occurs when the connector is used to link a business to a framework. A great number of security and homogenization services are performed inside the connections just before to the data transmission. The actual large data storage systems are mounted atop the framework-side connection after they have been assembled.

- **Big data storage and computation platform:**

The platform for the storage and computation of big data is nothing more than a collection of traditional storage and processing devices that are linked together through a network and set up in cluster mode for the purpose of sharing resources and improving performance. An open source Apache Hadoop Ecosystem serves as the basis for the configuration of the cluster. The network of open source components that is Apache Hadoop is transforming the way in which enterprises store data, manage data, and analyze data. In contrast to more conventional systems, Hadoop makes it possible for several analytical processes to be carried out concurrently on the same data while leveraging technology that is freely accessible to the public. Installing the Apache Hadoop Ecosystem requires the usage of Cloudera's open source platform, which is known as the CDH distribution.

- **Data management:**

This layer is used to organize and model the data stored in the big data infrastructure. This layer will be also responsible for handling any data privacy issues and handling the collaboration aspects of the logistics. The following services are provided:

Annotation/Curation Service: The data that will be used for the scenarios will have categories assigned to it, and annotations will be added to them. The service will be created on top of the Hadoop Cluster using the required tools based on the particular Hadoop distribution that is currently available. The relevant information, including description of the dataset, use of the dataset, version control, and so on, will be retained in a methodical way for each dataset, regardless of whether the dataset was imported in its raw format or in its processed format.

Data engineering: Each data collection will get descriptive analytics, often known as data visualization, courtesy of the data engineering service. In addition to this, it will take care of the process of data homogenization by converting each data component to the appropriate equivalent form that is suggested by the data model. In addition, the data will be cleaned up and modified such that any instances of null values that can be found will be removed. Additionally, the data will be prepared for the service in the appropriate forms, which are necessary for higher level analytical services. Python, Apache Spark, or both will be used throughout the development of this service. In order to overcome the issues posed by big data, these technologies were chosen because of their adaptability and capacity for growth within a distributed architecture.

Data modelling service: In order to roll out this service, it is necessary to construct a suitable data model by drawing inspiration from the many use cases and the data that is already at hand. In order to make the data model usable by a wide variety of data sources, it is necessary for each element of the model to have its own distinct mapping. The data model will be installed into one of the open source databases that are part of the infrastructure, and data will be imported for the purpose of being used by the analytics algorithms.

Big Data Business analytics: This layer includes all the business analytics required for supporting the use case scenarios.

Descriptive analytics and real-time analysis: To enable information interpretation and structuring, the discovery of hidden patterns and correlations, the induction of knowledge, and the creation of learning systems, a component of analytics operating in close proximity to real time is necessary. Because of its adaptability to the design of processes that can manage the complexity of commercial and scientific applications, new Semantic Models, and the quick development of cloud services to create new capabilities, it has a value that cannot be matched by any other technology. The ease with which data scientists (who work with algorithms and data transformations via a visual interface) and software engineers (who work with the concept of services to be invoked) can collaborate to reduce "time to market" in the development of complex big data and artificial intelligence applications is another notable feature of the Moriarty framework. This feature is particularly useful for accelerating the process of bringing new products to market.

Sentiment analysis:a component for analyzing sentiments and compiling statistics based on information obtained from social media. A crawler that scours e-commerce websites in search of customer reviews, performs an analysis of user ratings, and stores the results in documents that have been preformatted with properties such as: shop name, shop total votes, review date, positive or negative key features (selected from preset lists), review main text, and review star rating. The comments have been organized into favorable and negative categories, and the words that appear most often in both sets of assessments have been highlighted. In addition to this, it labels the various parts of speech and keeps only the nouns, adjectives, and adverbs that are associated with an emotion (consumer insights). Finally, it classifies user evaluations according to the frequency with which these consumer insights are spoken. The words are then linked using technology that is used for text mining in order to have a better understanding of how the previously described notions may represent the behavior of consumers.

Dynamic Routing: This component generates the best viable pathways for both direct and reverse flows in order to simplify the distribution operations of a variety of enterprises and enable them to share common logistics. This component offers solutions for the vehicle routing problem (VRP) as well as the vehicle routing problem with time windows (VRPTW) so that this objective may be accomplished. The Vehicle Routing issue (VRP) is a combinatorial optimization and integer programming issue that aims to give the best practicable set of routes for a fleet of vehicles to travel in order to transport things to a specific group of consumers. Specifically, the goal of this problem is to minimize the total amount of time and distance that the cars must travel. The overall route expenses are something that the VRP hopes to bring down. Heuristics and genetic algorithms are used because of the sheer number and frequency of real-world VRPs that they are required to solve. The VRPTW is the same issue as the VRP, with the additional constraint that, with the VRPTW, a time frame is associated with each customer,

ISSN: 2366-1313

giving an interval where the customer needs to be provided. The VRPTW is essentially the same as the VRP. The situation involves the delivery of items that are held at a central depot to clients who have made orders for such things and are waiting for delivery. The time period in question is what people at the depot describe to as the scheduling horizon.

Deliveries and demand forecasting: This component is responsible for configuring and using various forecasting algorithms in order to predict potential delivery issues in the future. There are a variety of ways for making forecasts, including statistical, economic, and even "soft computing" approaches. The context of the forecast, the relevance and availability of past data, the degree of accuracy desired, the time period to be forecasted, the cost/benefit (or value) of the prediction to the company, or the length of time available for the research will all have an impact on the methodology that is used. For the sake of forecasting, a behavior of dynamic method selection will be constructed.

Location optimization: This component will provide a number of options for analyzing the partners' logistics network architecture and identifying the most advantageous spots for click-and-collect stations and micro hubs. It enables an established and transparent logistical decision-making assistance, in addition to providing a wide range of features, such as extrapolation of cargo volumes, identification of optimal warehouse locations, route planning, and so on. In addition, it determines the most efficient logistics sites by taking into account the quantity of the cargo, the price data, and the distances involved. Transportation is meticulously tracked all the way down to the stock keeping unit (SKU) level, and optimization is performed based on actual freight charges. There is a possibility that overall transportation costs will go down as a result of improved network topologies.

An ideal network strategy for the placement and design of micro hubs will be devised, and it will take into account both the existing structures and the multistage logistics networks.

Inventory Routing: In the context of Vendor Managed Inventory (VMI), one of the more recent challenges in the field of operational research is known as the Inventory Routing Problem (IRP). A provider ships items to clients who are located in different parts of the world, according to the conventional representation of the issue. The supplier is responsible for controlling both their inventory and their stock levels, as well as preventing any of its customers from experiencing stock outs. As a direct result of this, the supplier is responsible for determining when to replace inventory based on the level of demand and the amount of product that is now available.

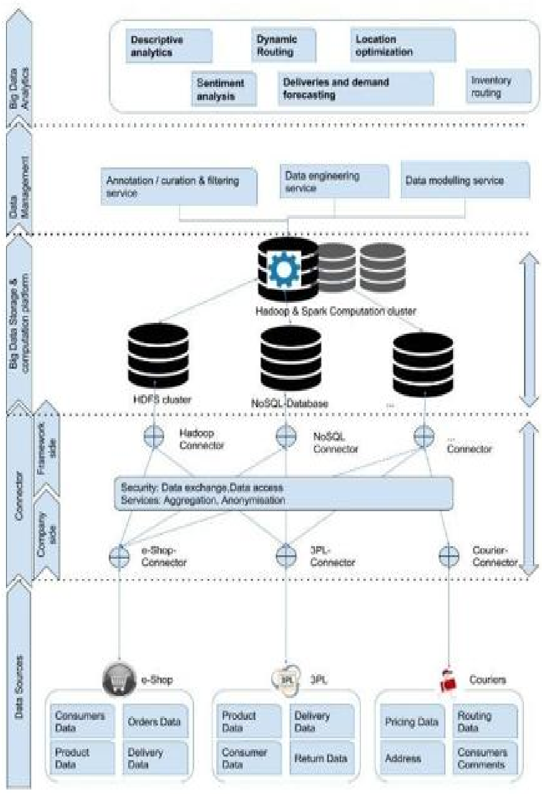

Figure 2- Overall architecture

Every single customer needs to have their access limited. As a direct result of this, it is not feasible to analyze all of the potential solutions and decide which one is the best using computational means. A problem with this kind of structure is said to be NP-complete. A customized algorithm that examines the data provided to determine, based on the results of the analysis, whether or not it would be more efficient to employ the whole supermarket network rather than just the central hub. It is possible to make an estimate of the cost savings that are anticipated, and the level of success may be evaluated using additional measures such as vehicle and fleet usage.

6. CONCLUDING

This research delineates a range of operational prerequisites that are characteristic of the e-commerce logistics sector of the economy. Then, it demonstrates how big data analytics may help with data integration, decision-making, and the execution of seamless logistical solutions to fulfill the needs that have been set for the organization. It is recommended to have an illustrative big data analytics architecture. Using sophisticated analytics will be the focus of the subsequent phase, which will see the suggested architecture and its associated use cases put into action. The difficulties of implementation will be examined, with a focus placed on demonstrating the ways in which big data analytics has changed the logistics of online shopping.

REFERENCES

[1] Addo-Tenkorang, R, and P. T. Helo (2016). Big data applications in operations/supply-chain manag-ement: A literature review. Computers & Industrial Engineering 101 (2016) 528–543.

[2] Agatz, N. A., M. Fleischmann and J. A. Van Nunen (2008). E-fulfillment and multi-channel distribution–A review. European Journal of Operational Research 187(2): 339-356.

[3] Akter, S. and Wamba, S.F. (2016). Big data analytics in E-commerce. A systematic review and agenda for future research. Electron Markets, 26, 173-194, 2016.

[4] Aziza, B. (2017). Amazon Buys Whole Foods. Now What? The Story Behind The Story. Available from https:// www. forbes.com/sites/ciocentral/2017/06 /23/amazon-buys-whole-foods-now-what-the-story- behind-the-story/#26fb066ee898

[5] Bask, Lipponen, and Tinnilä, (2012). E-Commerce Logistics: A Literature Research Review and Topics for Future Research. International Journal of E-Services and Mobile Applications, 4(3), 1-22, July-September 2012.

[6] Berente, N., and K. Lyytinen (2006). The Iterating Artifact As a Fundamental Construct in Information System Design, In Proceedings DESRIST, Claremont, CA: Claremont Graduate University.

[7] Boyer, K. K., Prud'homme, A. M., and Chung, W. (2009). The Last Mile Challenge: Evaluating the Effects of Customer Density and Delivery Window Patterns. Journal of Business Logistics, 30(1), 185–201. https://doi.org/10.1002/j.2158-1592.2009.tb00104.x

[8] Castiñeira, R, and Metzger, A. (2018).The TransformingTransport Project – Mobility Meets Big Data. In Proceedings of 7th Transport Research Arena TRA 2018, April 16-19, 2018, Vienna, Austria

[9] Chen, H., Chiang, R. H. L., and Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. MIS

Quarterly, 36(4),1165–1188.
[10] Danloup et al., 2014. Improving sustainability performance through collaborative food distribution International Conference on Green Supply Chain, France.
[11] Du, T. C., Li, E. Y., and Chou, D.(2005). Dynamic vehicle routing for online B2C delivery. The international journal of Management Science (33), 33–45. doi:10.1016/j.omega.2004.03.005
[12] E-commerce Europe. (2016). European B2C E-commerce Report - Facts, Figures, Infographic & Trends of 2015 and the 2016 Forecast of the European B2C E-commerce Market of Goods and Services. Retrieved from https://www.ecommerce-europe.eu/app/uploads/2016/07/European-B2C-E-commerce-Report-2016-Light- Version-FINAL.pdf
[13] Esper, T. L., T. D. Jensen, F. L. Turnipseed and S. Burton (2003). "The last mile: an examination of effects of online retail delivery strategies on consumers." Journal of Business Logistics 24(2): 177-203.
[14] Fabian, J. and Christian, D. (2012). Vehicle routing for attended home delivery in city logistics. Procedia: Social and Behavioural Sciences (39), 622–632. doi: 10. 1016/ j.sbspro. 2012. 03.135