DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF SOCIAL SCIENCES
DEPARTMENT OF BUSINESS ADMINISTRATION
BUSINESS ADMINISTRATION PROGRAM
MASTER'S THESIS

MATERIAL DATA FLOW IN LOGISTICS BASED ON GRAPH DATABASE

Yeliz Alnıaçık

Supervisor
Assoc. Prof. Dr. Aşkın ÖZDAĞOĞLU

İZMİR - 2021

APPROVAL PAGE

DECLARATION

I hereby declare that this master's thesis titled as **"Material Data Flow in Logistics Based on Graph Database"** has been written by myself in accordance with the academic rules and ethical conduct. I also declare that all materials benefited in this thesis consist of the mentioned resources in the reference list. I verify all these with my honour.

Date
26/07/2021
Yeliz ALNIACIK

ABSTRACT

Master's Thesis

Material Data Flow in Logistics Based on Graph Database

Yeliz ALNIAÇIK

Dokuz Eylül University
Graduate School of Social Sciences
Department of Business Administration
Business Administration Program

Transport management and network design strategies have a huge and significant impact on decision making process of global manufacturers due to their intensive logistics activities. In this study, one of the biggest producers of industrial and automotive products will be examined regarding its material flow in worldwide outbound shipments. The aim of the study is increasing transparency, visualize worldwide outbound transports, understanding and monitoring material flow and with including performance quality of international forwarders based on graph data to find better optimization solutions for future logistics network analysis. On this way, integrated common data platform will be created and one of the problems will be solved in logistics network design department. A modern approach will be defined with graph database and big data management will be analyzed and observed in structured logistics' outbound data.

Methodology is modelling relational database from SAP and ensuring data migration to graph database to provide pellucidity on outbound shipments for one-year logistics' data. Data is modelled with an Entity Relationship Diagram and collected by using SAP (EP1) and Microsoft Structured Query Language (MSQL) Server in relational database.

Keywords: Big Data, Data Visualization, Graph Database, Network Design, Network Management, Relational Database, Optimization in Logistics, Transport Management.

ÖZET

Yüksek Lisans Tezi

Grafik Veri Tabanına Göre Lojistikte Materyal Akışı

Yeliz ALNIAÇIK

Dokuz Eylül Üniversitesi

Sosyal Bilimler Enstitüsü

İngilizce İşletme Anabilim Dalı

İngilizce İşletme Yönetimi Programı

Taşıma yönetimi ve ağ tasarımı stratejileri, yoğun lojistik faaliyetleri nedeniyle küresel üreticilerin karar verme süreçlerinde büyük ve önemli bir etkiye sahiptir. Bu çalışmada sanayi ve otomotiv ürünlerinin en büyük üreticilerinden olan bir şirketin dünya çapındaki sevkiyatlarındaki malzeme akışı incelenecektir. Çalışmanın amacı, grafik veri tabanına dayalı şeffaflığı artırmak, dünya çapındaki gönderilen yüklemeleri görselleştirmek, malzeme akışını anlamak, izlemek ve gelecekteki lojistik ağ analizi için daha iyi optimizasyon çözümleri bulmaktır. Bu sayede bütünleşmiş ortak veri platformu oluşturulacak ve lojistik departmanındaki sorunlardan biri çözülecektir. Graf veritabanı ile modern bir yaklaşım tanımlanacak ve yapılandırılmış lojistiğin yüklemelerinde büyük veri yönetimi analiz edilecek ve gözlemlenecektir.

Bu çalışmanın yöntemi, SAP'den alınan verilerle ilişkisel veri tabanında modelleme yapmak ve bir yıl içerinde yüklenen gönderilerde netlik sağlamak için grafik veri tabanına veri geçişini sağlamaktır. Veriler, varlık bağlantı şeması ile modellenip, SAP (EP1) ve ilişkisel veri tabanının bir ürünü olan Microsoft Yapılandırılmış Sorgu Dili veri tabanı yönetim sisteminin yardımıyla toplanır.

Anahtar Kelimeler: Ağ Tasarımı, Ağ Yönetimi, Büyük Veri, Grafik Veri Tabanı, İlişkisel Veri Tabanı, Lojistikte Optimizasyon, Taşımacılık Yönetimi, Veri Görselleştirme.

MATARIAL DATA FLOW IN LOGISTICS BASED ON GRAPH DATABASE

CONTENTS

CHAPTER TWO

GENERAL INFORMATION ABOUT COMPANY & GRAPH DATABASE

CHAPTER THREE
ANALYSIS AND RESULTS

ABBREVIATIONS

1PL	First Party Logistics
2PL	Second Party Logistics
3PL	Third Party Logistics
4PL	Fourth Party Logistics
5PL	Fifth Party Logistics
AGV	Automated Guided Vehicles
BD	Big Data
CA	Continuous Approximation
CEO	Chief Executive Officer
CPU	Central Processing Unit
DBMS	Database Management System
ERD	Entity Relationship Diagram
ETL	Extract, Transform, Load
GDBMS	Extract, Transform, Load
GIS	Graph Database Management System
HTML	Hypertext Markup Language
IBM	International Business Machines
ICC	International Chamber of Commerce
IT	Information Technologies
IoT	Internet of Things
ITS	Intelligent Transport Systems
JSON	JavaScript Object Notation
KPI	Key Performance Indicator
LPI	Logistics Performance Internal
LPE	Logistics Performance External
LPF	Logistics Performance of Forwarder
LSP	Logistics Service Provider
MATLAB	Matrix Laboratory
MSQL	Microsoft Structured Query Language

OEM	Original Equipment Manufacturer
RAM	Random Access Memory
RFID	Radio Frequency Identification Devices
RPA	Robotic Process Automation
SAP	Systems, Application & Products
SCM	Supply Chain Management
SSSP	Single Source Shortest Path
VRP	Vehicle Routing Problem

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDIX

INTRODUCTION

In intensive international trading activities, logistics operations play a significant role. The effective and efficient logistics operations create value added services and coordinated flow of information, products and cash within and across and national borders under the impact of globalization. However, generating this frame based on efficient and effective logistics operations is not simple. Most of the time managing and monitoring highly volumed logistics activities are quite hard due to diversified and huge amount of data which obtain from these activities. Especially multinational companies who operates overseas market areas suffers on logistics' data analysis. This research focuses on one of these companies: Schaeffler Technologies AG & Co. KG. which is a global supplier based on automotive and industrial sectors and its inbound and outbound shipments all over the world. The company covers European and all major international markets with its production and logistics activities. This study aims to to understand material flow in the company and solve its unneccesarry logistics activities which increase cost and time consuming in outbound shipments.

Today, logistics' data analysis is made with traditional databases such as relational database. However, new approaches are rising for databases like graph database. There is a need for storing highly volumed data, analyzing, and visualizing it. Generally, these desires can be made in separated platforms. Nevertheless, graph database is giving an opportunity to store, analyze and visualize the data at the same time with ensuring highly retrieving data performance. That's why, given possibilities of graph database will be used in logistics' data of Schaeffler Technologies AG & Co. KG. to analyze and find better optimization solutions within current logistics network structure. Methodology is data migration from relational database to graph database and data modelling an Entity Relationship Diagram (ERD).

During this study, it was a pleasure to work with Mr. Arash Attarzadeh, many thanks to him for technical support with his deep knowledge about data architecting and data management systems as a colleague in department and to Mehmet Iriz due to his approaches and leading the way of project path as a department manager.

In the first chapter, logistics operations and management will be clarified regarding the literature review based on determined concepts. On the other hand, the second chapter is going to provide general information about the company and detailed information about graph database. In following third chapter is going to identify methodology of this study and how data migration is occurred and data model is created after all research questions will be answered, analyzed and shown in a practical way.

.

CHAPTER ONE
LOGISTICS OPERATIONS AND MANAGEMENT

In the first chapter, the concepts of logistics, importance of efficient logistics activities and challenges which are getting faced during the logistics operations will be examined in detail.

1.1. THE CONCEPTS OF LOGISTICS

Logistics is accepted as a crucial function within business by most of logistics expert and academician (Rutner & Langley, 2000). It is a common clue that logistics involves the flow of physical goods from one place to another.

There are many ways to define logistics, one smooth definition is the "Seven R`s of Logistics". Rutner & Langley (2000: 73) refers Seven R's as;

"providing the existence of right product, in the right quantity and the right condition, at the right place, at the right time, for the right customer, at the right cost".

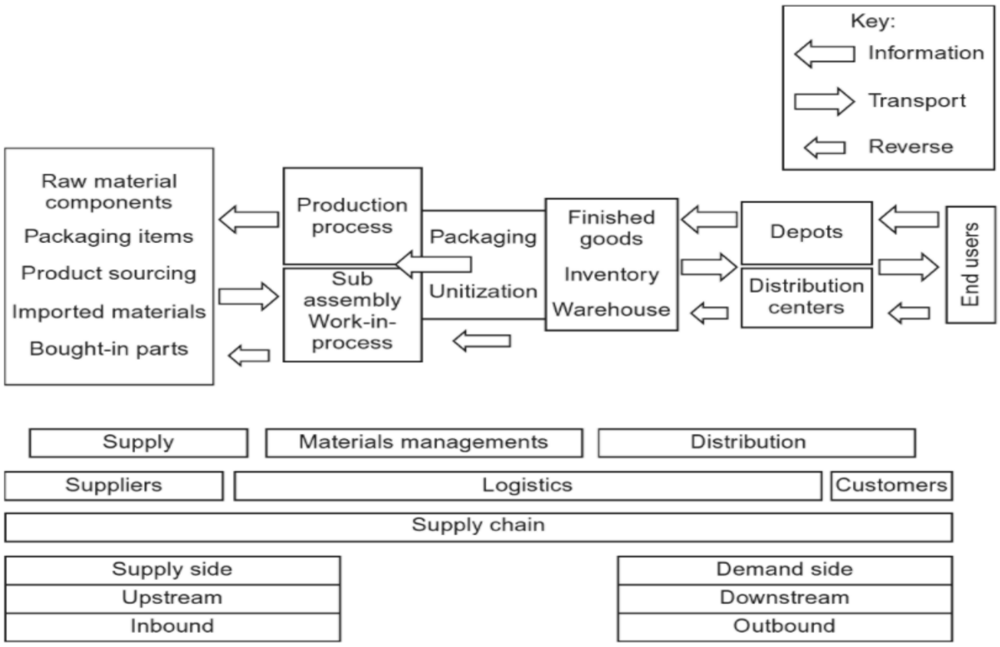
In 1947, Simpson and Weiner mentioned about logistics term as Americans use in a meaning of the technique of packing stores which comes from the French word; "maître du logie" (Lummus et al., 2001). Logistics cannot be examined apart from carried materials or products. Cavinato (1982), determined logistics as the management of all incoming and outgoing materials, items, provides and finished goods (Lummus et al., 2001: 1).

On the other hand, Council of Logistics Management (1998) indicated as the progress of scheduling, applying and managing the effectual, efficient flow and warehousing of goods, services and relevant information from the spot of beginning to the point of consumption in the frame of accepted customer requirements (Lummus et al., 2001: 1). Customer service and customer satisfaction are identification of logistics in the literature review (Rutner & Langley, 2000: 1). In order to ensure customer service and customer satisfaction salutarily, the logistics process should continue under the highlight of a value chain. The value-added exchange operation can be short run for only one transaction or long run in a covenanted relationship between buyer and seller (Rutner & Langley, 2000: 2). In any type of contractual relationships, the value-added logistics activities should be done adequately.

The literature review in the search of logistics concept showed that there are many studies based on differences between Supply Chain Management (SCM) and logistics as a term. In order to clarify this case, in this paper as well; SCM contains the flow of logistics, production progress, transaction of information, observing of all actions during the supply chain process (Lummus et al., 2001: 6). This explanation indicates that logistics activities are part of the supply chain process. Since, SCM covers also innovation and product development, purchasing and sourcing, manufacturing and operations, customer service, global deployment procedures. In the variant of logistics accepted as the administrative science relevant to supervising the whole "supply chain", it is a similar metaphor which means encircle the "life span" of a goods, from descent raw material to consuming and disposal by the final customer (Lecavalier, 2016).

Figure 1 shows that how supply chain process contains to the logistics activities inside of it. It is understandable that there are many significant logistics` terminologies which are going to be considered in the following parts of the study.

Figure 1: Logistics Flows and Some of The Different Logistics Terminologies



Source: Farahani et al., 2011

1.1.1. Transportation System

A real multinational company searches for settling business activities within and beyond of multiple nations through worldwide brands and integrated facilities (Bowersox & Calantone, 1998: 3). In this wide frame, transportation system carries an important role during logistics operations. Companies and their commodities` markets are usually detached geographically (Farahani et al., 2011: 13). The period and place utility of commodities by providing them at appropriate time and to correct location where they are required are increased by transportation (Farahani et al., 2011: 13). In order to reach level of customer satisfaction, mentioned essential point should be main consideration during transportation. However, logistics costs increase while providing the right time and the right place factors. The assessment of transportation systems includes the quality of transportation capabilities (Witte, 1980). Logistics cost differentiates according to transport modes, delivery term and reverse logistics concepts to make more qualified transportation surely regarding to bilateral agreement between buyer and seller.

1.1.1.1. Transport Modes

Minimizing overstocking, ensuring optimum quantity, delivering just-in-time are the main concerns in logistics activities. In order to ensure these desires, different types of transport modes are used by carrier for physical flow of goods.

Most of the global companies perform in local area by truck and in foreign markets by container in short-haulage shipment with diverse transport modes. (Fan et al., 2019: 1). However, there is no fixed transport mode for the global companies. Moreover, there might be intermodal freight to serve same logistics operation between sender and receiver. According to most of research paper, basically there are five types of transport modes. These are sea, rail, road, air and pipeline transports.

Road Transport: It is one of the most preferred transport modes in the logistics activities. The main benefits of road transport in comparison with other modes of transportation are elasticity and adaptability (Farahani et al., 2011: 13). Door-to-door services is possible via trucks between start point and end point without either loading or unloading and it gives another level flexibility to roadway transportations by largest scope of vehicle forms for products almost any volume and amount (Farahani et al., 2011: 13). Also, the ratio of damages and loss of products are rather greater than air freight and less than railway transports.

Rail Transport: The second mostly chosen transport mode is rail transportation. Rail transport is common between major cities in the world due to infrastructure manner and designed railway network channels. It creates inflexible transportation because of existence of terminals in origin and destination place.

Sea Transport: This carriage is known as the oldest transport mode. It gives opportunity to ensure the cheaper transport rather than other transportation modes. However, speed of the transport decreases. The products which are not level of emergency and huge sizes might carry by sea freight regarding location of the sender and receiver. It is basically preferred in bulky and heavy commodities by the companies.

Air Transport: It provides a service based on low transit time and high transportation cost. It is mostly selected in premium urgent shipments. Nevertheless, the costly air transport might be considered under the trade off perspective due to cut down on warehousing and inventory costs; there might be specific situations for high quality goods, perishable products, limited trading periods and emergent cases (Farahani et al., 2011: 15). The loss and damages of commodities are relatively less than other transport modes. Also, regarding these situations, companies prefer whether air transport or another transportation mode.

Pipeline Transport: It is preferable for the long distance and low-cost carriage. Basically, pipeline systems used for specific product such as refined petroleum, natural gas, chemicals, oil, water and some other product type which are not solid. Reliability is quite important while using pumping equipment. Pipeline transport is not open to environmental effects of logistics condition. Transportation cost basically relies on the volume of the product to carry.

On the other hand, different transport modes might have interaction within same network. So, a company does not have to select only one transport mode. On this case multimodal or intermodal transportation should be under the consideration. The tendency on the volume of vehicles rises meanwhile individual shipment diminishes means that a higher number of dispatches must be consolidated in every intermodal charging set (Vasiliauskas, 2010). Vasiliauskas (2010) mentions about intermodal transportation is a necessity to aggregate products flow with the purpose of fast and cheap transshipments between modes of transportation. Also, multimodal transportation carries an important role in logistics activities. United Nations Economic Commission for Europe (2009) defined international multimodal transportation as freight for cargo using two or more various transport modes (Kim et al., 2020). In environmental aspect, logistics activities and supply chain management create problems in the world’s different regions. Environmental management might be crucial in its proactive status in a frame of organizing multimodal transportation centers where integrate various types transport companies, logistics services and producing or progressing activities at the unique point (Kengpol et al., 2014). Multimodal transportation quite beneficial for transit time, optimized cost and CO_2 emission aspects. Although intermodal and multimodal transportation look like so similar, there are some differences between those two concepts. In multimodal transportation only one transport contract is made from loading point until unloading location between a firm and logistics company. However, different carriage contract is signed for every transport mode during the logistics activities between a firm and agreed forwarder. Only one forwarder is responsible and managing logistics activities during the carriage in multimodal transportation, unlike intermodal transportation. In intermodal transportation, there are more than one carrier is liable on carriage. On the other hand, all transport mode might be combined in multimodal transportation, but only roadway, sea freight and railway can be combined in intermodal transportation.

The decision regarding the selection of the transport mode is quite complicated due to the volume of consolidation together with the many approaches of examination and valuation of each selection (Slater, 1979). For this reason, companies need to evaluate cost and time efficiency simultaneously and adequately for the choice of transportation modes.

1.1.1.2. The Delivery Term (Incoterms 2020)

The delivery terms indicate which side is covering cost, risk and obligations according the determined agreement between buyer and seller in mostly international trading activities. The aim of the Incoterm is setting out the essential roles and duties of the purchaser and the vendor in a frame of delivery and risk based on surrounded export or import contract; it is also proper for domestic market (Committees, 2020). After 2010 Incoterms, International Chamber of Commerce (ICC) defined Incoterms 2020 rules according to new trends in logistics activities. These rules are examined by ICC in two categories. The first one is “Rules for Any Mode or Modes of Transport” and another one is “Rules for Sea and Inland Waterway Transport”. In a comparison between Incoterms 2010 and 2020; Delivery at Terminal (DAT) renamed as Delivered at Place Unloaded (DPU). The other Incoterms remained as the same.

Ex Works (EXW); Sender of the goods have no costs and risks in this delivery term. All responsibilities belong to receiver during the whole logistics activities. That’s why sender is not allowed to make any agreement with the logistics’ company. Also, importer is responsible all customs procedures in both sides.

Free Carrier (FCA); In this delivery term, sender and receiver determines a location where sender would deliver goods to receiver. Until this place, shipping and customs office procedures are belonging to sender. After delivery of the goods to receiver, transportation cost and responsibilities are passing to receiver. There is no any obligation for the insurance for the both parties. The delivery term might be used for every transport mode.

Carriage Paid To (CPT); According to agreed location where geographically closer to receiver in a comparison with FCA, goods are carried by sender and sender is liable of logistics’ cost and all other logistics activities up to determined location between sender and receiver. After the delivering of the goods to receiver, all risk and responsivities are transferred to receiver, except freight cost of the transportation.

Carriage and Insurance Paid To (CIP); This delivery term is quite similar to CPT delivery term, but there is a significant difference between them. Sender must deliver goods to receiver as all transport cost and insurance are already paid. The main distinction is insurance costs during the transportation between CPT and CIP.

Delivered at Place (DAP); It is one of the less risky delivery terms for the receiver. Sender takes all responsibility for transportation until the place where is close to receiver and this place might be even receivers' factory location.

Delivered at Place Unloaded (DPU); Loading and unloading cost belongs to sender where products are going to deliver receiver at agreed place. However, customs procedures belong to receiver in receivers` country.

Delivered Duty Paid (DDP); In this delivery term, receiver has no responsibility during transportation process. All customs, loading, unloading costs belong to sender nearby physical logistics cost.

All these mentioned delivery terms are "Rules for Any Mode or Modes of Transport". **Figure 2** shows all rules for any transport modes based on cost, risk and insurance criteria of senders and receivers' responsibilities.

Figure 2: Rules for Any Mode or Modes of Transport

Source: ICC National Committees, 2020

Rules for Sea and Inland Waterway Transport are as below as it is indicated on **Figure 3**;

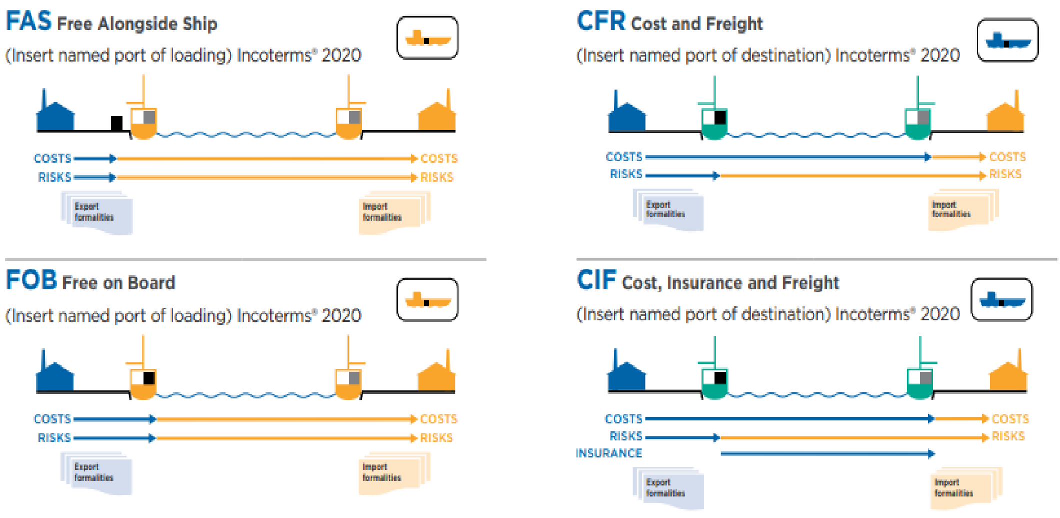
Free Alongside Ship (FAS); This delivery term is sea freight version of the FCA incoterm. Sender is responsible until the seaport to carry products. Up to specified location, sender is responsible on transportation and customs procedures in domestic country.

Free on Board (FOB); In an addition to FAS incoterm, sender carries product until seaport however, it is also responsible for the unloading on shipboard. After unloading the products to shipboard, all responsibilities, risks, customs procedures and freight cost are transferred to receiver.

Cost and Freight (CFR); This delivery term is sea freight version of the CPT incoterm. Sender has to take on all obligations and liabilities until port of destination. Although freight cost pertains to sender of the goods, risk of the carriage belongs to receiver.

Cost, Insurance and Freight (CIF); CIP and CIF are the mere delivery terms which includes insurance of the goods during the transportation. If the transportation will be made by sea freight and desired to be included insurance; CIF should be selected between parties.

Figure 3: Rules for Sea and Inland Waterway Transport



Source: ICC National Committees, 2020

1.1.1.3. Logistics Parties

Logistics parties indicates service integration during the logistics operations. The parties help to understand responsible organization of delivery and shipments with details. Due to importance of logistics concepts, 1st Party Logistics (1PL), Second Party Logistics (2PL), 3rd Party Logistics(3PL), 4th Party Logistics (4PL) and 5th Party Logistics (5PL) will be shortly mentioned. 1PL defined as when a firm ensures its logistics' activities itself and carry its goods and merchandises from start point to the end point with its own vehicles. Also, 2PL includes asset-based carries. It means that a company can work in a frame of leasing or determined contacts for airline or ships regarding different transport modes.

3PL involves diverse sort of services. Therefore, this service is counted as predominant concern in the transportation. Current business tendency shows the third-party logistics is provides a balance for considerable expansion (Sink et al., 1996). Virum (1993), submitted principal "Nordic" explanation about 3PL as;

"*The services offered by a middleman in the logistics channel that has specialized in providing, by contract, for a given time period, all or a considerable number of logistics activities for other firms*." (Stefansson, 2006: 80).

3PL comprises supply chain management, warehousing, logistics IT software opportunities, customs brokerage etc. So, 3PL service integration does not only consider on transport from one point to another point. Logistics Service Providers (LSPs) becomes outsource for a firm during the whole process of logistics activities. On this way, several actors play different and significant roles in integrated transportation services. Frequently these services were performed in a frame of a "healthy" competition. Therefore, the parties had "arm's length relationships" in order to keep information flow in minimum level (Skjoett-Larsen, 1999). However, today logistics parties aim to ensure flexibility and good service for their customers based on specified requirements.

Many of 4PL companies ensures services to customer with their duty and know-how to fulfill customers' regulations instead of having stockpile facilities or fleet etc. (Stefansson, 2006: 81). The physical transport of the merchandises is contracted out to other third-party service providers (Stefansson, 2006: 81).

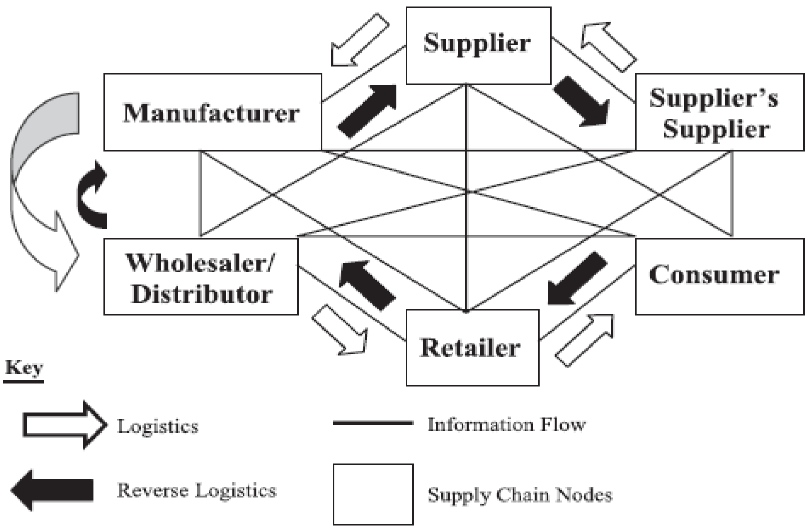
On the other hand, 5PL logistics provider termed as a logistics collector, without any asset. The demand is consolidated with the demand of 3PL into other bulk capacity to get better rates with diverse types of airlines and shipping forwarders.

1.1.1.4. Reverse Logistics

Reverse logistics simply defines as returning goods from end customer to original producer. Lu and Bostel (2007) examined reverse logistics arise from a series of activities from the progress of returning a merchandise to that merchandise being fully fixed or truly discarded (Chiu et al., 2011). So, status of the goods is quite significant due to potential of the re-manufacturing of that product. Therefore, returned goods must be dismantled, controlled and discarded (Chiu et al., 2011: 513).

Reverse logistics involves a series of operations from resending back a merchandise. **Figure 4** shows how is the reverse logistics flow look like between parties. It starts from consumer and follows a way until first supplier of the merchandise. Thus, reverse logistics creates extra logistics cost and increases time consumption.

Figure 4: Consumer Supply Chain

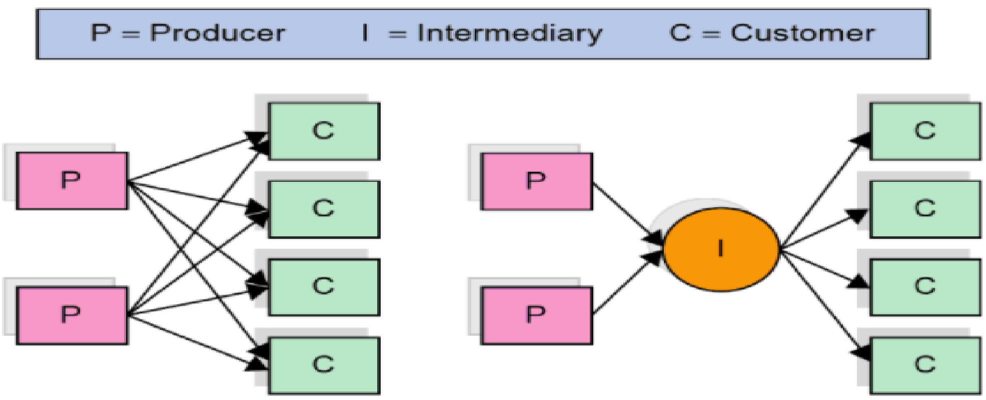


Source: Lummus et al., 2001: 6

1.1.2. Channels of Distribution

In order to find alternative options and path to send products to its marketplace, merchandises' distribution channel is one of the crucial and mostly challenging logistics concepts (Farahani et al., 2011: 26). As a management action, previously notion of logistics concentrated on circulation of goods and it became a business strategy to ensure advantages on time and location (Gundlach et al., 2006: 432). For this reason, specialized distribution activities play a fundamental role in logistics activities. It leads to increasing on intermediaries which are dedicated about those activities rather than polyvalent distributors, so many actors are involved until products are delivered to final customer (Gripsrud et al., 2006: 643). Most of the time, these actors called as intermediaries as also are shown in Figure 4. Some distribution activities include intermediaries, some of them occurs with direct shipments. These intermediaries might be retailer, wholesalers etc. Merchandises are sent to directly to end customer or firstly to a hub and then final customer by producer. Those hubs can be an outlet or warehouse of producer, customer or retailer as well in the process of physical transferring. **Figure 5** also indicates business ties between producer and customer. In first column, the number of links are counted as "P*C= 2*4=8" without intermediaries whereas in second column "P+C=2+4=6" with intermediaries. On this way, trading ties decreases with support of intermediaries. Distribution systems with intermediaries helps especially on merchandises which has large volumes.

Figure 5: Distribution Systems with and without Intermediaries



Source: Farahani et al., 2011: 27

The decision on changes about distribution channels is quite hard due to long-run agreements with intermediaries or various distribution actors in the system. There are some indicators in the distribution channels` system.

"Main drivers are "demand level", "service level", "product characteristics", "logistics costs", "labour and land", "accessibility" and "contextual factors" during the distribution channels decision making process." (Onstein et al., 2019: 243). Those mentioned indicators help to decision makers about how to decide on distribution channels during the logistics process.

The products might be distributed in "Centralized" or "Decentralized" structures. In general perspective "Centralized" distribution form means using an only distribution center location whereas "Decentralized" distribution structure consists of several dispersion unit location (Onstein, et al., 2019: 245). According to Kuipers and Eenhuizen (2009), six different distribution structures are possible, these structures are as below for the channels of distribution in a detailed approach:

Direct system is a method of sending the goods directly to the customers without any dispersion unit location.

Classical system includes one national shipping point and the products are distributed to customers from that determined location.

Multi-country system ensures a support for international distribution option. That's why in this distribution location companies are used to an international stockholding points in order to give an answer to their customers` demands.

Multi-country system II is a combination of Classical and Multi-country System I. Since, both of national and international shipping points assists as distribution locations during the logistics activities.

Cross-docking is another logistics strategy which is preferred by the companies when the material flow is under the prioritized consideration in the transportation network. Cross-docking strategy is basically described as the products are carried from providers to end user over shipping points without placing them within long run warehousing (Kücükoglu & Öztürk, 2017).

Cross-docking II system has a small difference from cross-docking due to additional stockholding point in the last stage before the items arriving to final customer. Except mentioned dissimilarity, all process is the same with cross-docking during the logistics activities.

1.1.3. Warehousing and Storage

During the intensive physical flow in logistics, warehousing and storage carry a significant role due to replacements of the products in the network bewteen buyer and seller. Especially in 3PL services, warehousing and storage ensures flexibility in time and volume level. Bartholdi and Hackman (2016: 3) refers warehouses as a place which goods are stopped and handled (Kembro et al., 2018: 892).

In globalized logistics' network area, warehousing and storage opportunities are used in order to decrease trasnport costs. The warehouse is also defined as an activity which is reducing the waste of non-value adding operations by a responsible party whereas enhancing effectiviness of the system (Abushaikha et al., 2018). In the efficient logistics activities, warehousing and storage are crutial ways to create adabtable lead times during the logistics operations meanwhile providing decreased inventory especially storage yards in the plant. Customarily, inventories are stocked in the plant`s warehouses. Nevertheless, lately manufacturers are concentrated their main producing activities and started to outsource warehousing and storage activities (Lee & Elsayed, 2005). It leaded to increasing scope of warehousing and storage usage.

In general aspect, warehousing and storage activities supports logistics parties based on handling, packing, storing area at the same time they present consoludated and centralized network and also flexible distribution opportunities.

Fan et al. (2019) shows there are significant operational difficulties and enormous opportunities in order to develop logistics activities. Therefore, importance and challenges in the logistics activities are going to be examined in following part of this paper.

1.2. IMPORTANCE OF EFFICIENT LOGISTICS ACTIVITIES

Magnitude of logistics activities are depending on early ages for every society. Man was considered about how to transfer materials to a construction side while building the great pyramids anciently (Lummus et al., 2001: 1).

In early referrals, logistics are mentioned as primarily activities also in military. Logistics operations carried an important role in movement of troops and supplying process of ammunition since early ages. During both World Wars, logistics took great attention from armed forces (Lummus et al., 2001: 1). It brought a necessity for great bargain for planning and mathematics as much as logistical physical activities.

Today, logistics operations accessed another level of importance all over the world with the huge impact of globalization. Globalization obliges a strongly coordinated flow of information, products and cash within and across and national borders (Bowersox & Calantone, 1998). Eventually, the significance of logistics consists the capability to figure out storage, transportation, packaging problems, meanwhile the rivalry is rising in the business world by expansion the country (Marti et al., 2014). On the other hand, bringing some solutions for the problems are not enough and sustainable during logistics operations. In general aspect, logistics' parties need to consider on efficient logistics process in order to enhance customer satisfaction. Thus, effectiveness and flexibility in the logistics system might be a perquisite to meet with definite demand of the customer (Klein & Morschett, 2006).

Multinational companies conduct with their rising number in local either foreign markets regarding various logistics exigences and operational rules in the concurrent globalized economy (Fan et al., 2019: 1). It also leads to increase importance of the transport management strategies of the global companies and logistics activities play a key role as part of supply chain management. Since, proper decisions for most of the logistics activities provide higher efficiency and lower costs for the company.

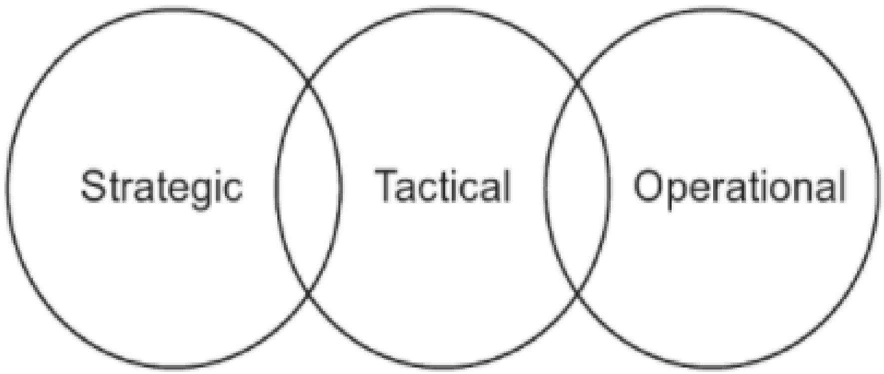
In the global trading area, efficient logistics activities are directly related with performance of the physical flow during the logistics operations. To understand the logistics operations are efficient or not, it is important to focus on the logistics performance and effectual logistics decisions.

1.2.1. Logistics Decisions

The smooth logistics operations are up to logistics decisions which are effective and efficient. There is a distinguished definition for effectiveness and efficiency in the logistics activities. Effectiveness is the orbit a target which has already been accomplished whereas efficiency is the level of consumed sources economically (Chow et al., 1994). So, logistics decisions need to be a frame of doing right things and doing things in a right way based on basic explanations of effectiveness and efficiency.

In a deeper explanation, logistics decisions are not coming from one unique basement. As it is shown in **Figure 6**, logistics decisions are interrelated in strategic, tactical and operational aspects. In the implementation process of logistics actions, value added activities are operational; planned and controlled transportations are tactical; disseminations of network design are known as strategic decisions (Hsiao, et al., 2010).

Figure 6: How Logistics Decisions Are Interrelated



Source: Farahani et al., 2011: 4

1.2.1.1. Strategic Decisions (Logistics Network Design)

Strategic decisions include long term goals in the logistics operations. Those decisions are taken in administrative level. There are three main objectives of the road map of logistics network design; these are capital deduction, reducing cost and customer satisfaction (SteadieSeifi, 2011: 47). In the purpose of making a decision for a long period of time, the data analysis quite significant via using current situation and predicting future logistics operations and demands on this way sustainability can be ensured in the logistics activities. According to SteadieSeifi (2011), below instruments are used commonly to provide mentioned objectives and sustainability.

Benchmarking is trying to find best option and standards regarding to comparative analysis. On this way, the gap between customers` desires and current logistics activities can be filled in performance and cost level.

Optimization Programming is another tool for strategic decision which applying heuristic algorithms, mathematical programming or making scenario analysis to find out better way for physical flow in logistics.

Continuous Approximation (CA) is a model which is used to achieve proximate optimum solutions under the large-scale problems during the logistics operations (Wang & Wu, 2019: 3). It is a helpful tool to show the most realistic result in the large amount of data for the strategic decisions.

Simulation assesses obedience of the system within various circumstances. It shows possible problems, helps to understand constraints in ongoing process of the strategy before spending capital.

Forecasting is the way of defining unclear variable based on available and current data source. It is basically used for short and medium run prediction of logistics activities which customers has certain demand in the future.

The strategic decisions illuminate the way of tactical and operational decisions within logistic activities. Since, it is required to improve method of operating and processes for the execution of the plans (Lambert & Stock, 1982: 40).

1.2.1.2. Tactical Decisions

Tactical decisions are related with how logistics activities should be planned and controlled regarding determined strategies. Tactical decisions are quite significant because envisaging and monitoring is a necessity within logistics network design. Otherwise, taken strategic decisions do not go with a swing and might affect logistics activities negatively. On the other hand, some researchers examine tactical decision as also appraisal decisions within the scheme of Supply Chain Network Design (SCND) and determine tactical decision as of short time impact of the strategic decisions (Nobari & Kheirkah, 2017: 410).

Qualified strategic and tactical decisions shows strength of the firm's marketing and supply chain strategies because there is a strong ties and commitment between those decisions and strategies (Bell & Chen, 2015: 78) . That's why, strategic decisions are minor and insignificant without tactical decisions.

1.2.1.3. Operational Decisions

Operational decisions are substantial to create value added operations during the logistics activities. Deciding on optimum routes based on logistics cost and lead time regarding customer preferences are accordingly effective and efficient operational decisions (Holcomb et al., 2014: 524).

In 1994, Holmstrom and Aavikko supported that possibility of increasing inventory turnover meanwhile logistics cost was decreasing around 10% and handling cost was remaining the same by developing the operational decisions. It is obvious that operational decisions are crucial although these decisions are taken for short-time logistics activities.

Mentioned three kind of decisions; strategic, tactical and operational decisions are integrated to each other, for this reason they cannot be under the consideration separately for the systematic logistics operations. There should be follow up decisions making process in this framework.

In the next part of this study, challenges which are faced within logistics operation will be examined.

1.3. CHALLENGES IN LOGISTICS OPERATIONS

In the scope of increased logistics operations, challenges and problems which have to be figured out is also raising up. In the beginning of this paper, determined logistics concept as doing "right" activity might goes wrong. It leads to many problems and challenges between logistics parties. During the logistics operations, it is highly possible that getting face with several challenges either in local market or overseas areas. Logistics service includes increased difficulties as operator during main transportation and storage activities meanwhile providing value-added proceedings to tackle strict margins (Selviaridis & Norrman, 2015: 592). These difficulties might occur in numerous aspects in the global range.

Firstly, COVID-19/SARS-CoV-2 pandemic crisis created huge impact at many sectors in the world and on the transportation area. Natural disasters create many obstacles for logistics activities like hurricane, earthquake, and tsunami etc. Nevertheless, a pandemic has rather special significance for the logistics operations because there is no identified peculiar geographical location and certain time limitation (Ivanov & Das, 2020: 91). Due to risk of the epidemic, governmental restrictions which constraints trading activities raised up all over the world. Disease outbreaks caused deceleration in economies and transportation system for every transport mode (Loske, 2020: 2). Epidemic created a domino effect in the global logistics sector and thus, transport volume and the number of shipments decreased in international supply chain.

Political decisions are another aspect for the examining challenges in logistics operations. For instance, on 29th March 2019, the UK decided to leave EU and this situation created uncertainty and apprehension for the manufacturer which are centered in the UK (Li et al., 2019: 174). Brexit (British exit) sparked off a gap with unclear information flow about logistics operations which is relevant inbound and outbound shipments for the UK. Especially trading activities between EU and the UK have been affected. Since, cost of customs duty, transit time and imported product taxes became under the consideration (Nakamura et al., 2019: 982).

Logistics activities also under the impact of environmental factors. Amendments of demographic and socioeconomic structures in shipper or receiver country, adaptation to domestic requirements, fulfilling the custom regulations and bureaucracy are effective on the domain of control during the logistics operations (Sandberg, 2011: 8).

The scope of technological developments based on smart logistics ensures many benefits for the ongoing process of international trading activities. Smart logistics apply artificial intelligence and IoT (Internet of Things) in order to design machine loading, check manufacturing flow, arrange vehicle routing, plan dispatches and circulation of vehicles (Bag et al., 2020: 5). Technological improvements give opportunities to use some implementations to increase traceability in logistics activities, such as automated guided vehicles (AGV), global information systems (GIS), radio-frequency identifications devices (RFID) etc. However, technological know-how, financial and infrastructural investments are not distributed equally to each country for the use of this kind of technologies. In this perspective, there might be some challenges during the logistics operations for the logistics parties who cannot catch logistics trends simultaneously in technological level.

On the other hand, during the physical flow of materials some difficulties might occur unexpectedly. In operational approach, recently Suez Canal in Egypt clogged by gigantic container ship and blocked one of the busiest sea routes (BBC, 2021). Also, pirate attacks overseas shipments are another operational risk during the logistics activities.

It is possible to examine many challenges within flow of logistics like planning, warehousing, handling, consolidating, vehicle routing problems (VRP) etc. Regardless, focused logistics' challenges for this study will be considered in the following side of this chapter. These are big data analysis, material flow and data visualization in the logistics operations.

1.3.1. Big Data Analysis in Logistics

Big data (BD) is quite famous term, but there is no specified description for it. The storing and data analytics are two prevalent spots of big data (Hurtado et al., 2019: 839). International Business Machines (IBM) defines big data as merely beyond of concerning data volume; big data is a chance to discover insight of the data and context which is new and emergent to run business further nimble and to respond questions over and above of the limit which is already examined (Porche et al., 2014: 2).

Big data ensures great opportunities to bring meaningful information for instance the direction of United State National Intelligence describes big data as the providing of mass analytics inside and crosswise data to allow integrity of information (Porche et al., 2014: 7).

In technical level, it is possible to store a bunch of data in traditional database with a limitless way. Despite that highly volumed data which is gathered and distributed becomes harder to make data mining and merging efficiently in due time (Porche et al., 2014: 4). In the logistics sector, highly volumed data needed to be analyzed. However, questioning right way, obtaining appropriate information processing, getting confidential data are the major challenges for big data in logistics (Hurtado et al., 2019: 839). On the other hand, tracking and tracing shipments are quite hard for every transport mode within big data which has continuing streaming real time data.

In data processing approach, variety and velocity of data are two significant challenges for data analysis in logistics operations (AlShaer et al., 2019: 793). Since, data is gathering from miscellaneous data sources. Variety of data is relevant with structure of the data sources. For instance, data within relational data base quite structured whereas JavaScript Object Notation (JSON) and Hypertext Markup Language (HTML) are semi-structured and text is unformed (AlShaer et al., 2019: 793). It creates difficulties to make data analysis within logistics data and affects speed of data flow negatively in velocity aspect.

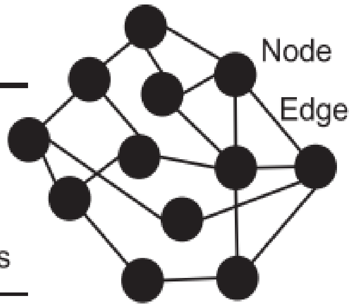
1.3.2. Material Flow in Logistics

Within physical flow in logistics, questioning essential information such as route planning, timing, appropriate transportation, requirements, competencies, tracking of needs and transparency are integrated each other (Chopra, 2020). In overall SCM process, material flow starts from original supplier when material is raw and the flow continues until final material arriving to end customer. This long progress creates complexity in networks of material flow. Due to mentioned complexity, in 2020 Funke

& Becker defines material flow in three perspectives. **Figure 7** shows network at global production, supply chain and material flow level and how would it look like in a graph. Global production network examined as relationship between firms for a particular material or services. Meanwhile, supply chain networks show transportation intercompany deliveries and material flow network indicates transaction of one product or numerous materials within factories.

Figure 7: Perspectives on Material Flow

Level	Node	Edge
Global Production Network	Company	Contracts
Supply Chain Network	Plants	Deliveries
Material Flow Network	Machines	Material Flows



Source: Funke & Becker, 2020: 297

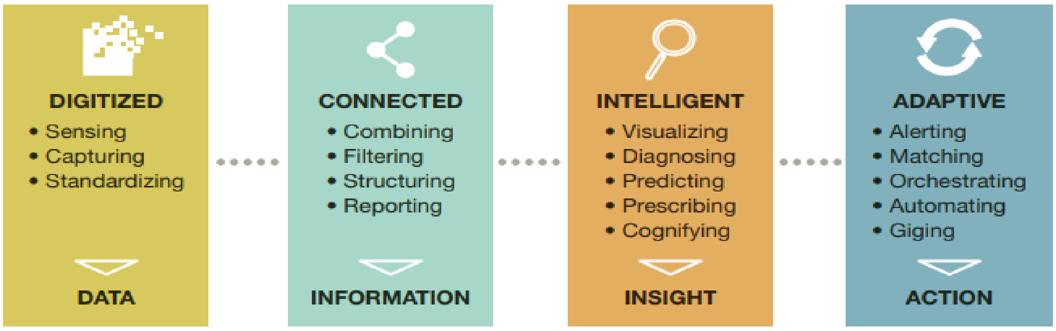
Physical distribution and logistics concentrate on coordination and functions between parties for product movements for a company, however they omit the cooperation within the company and exterior relationships with other channel parties (Hou et al., 2017: 107). It leads to lose material flow transparency during the logistics operations.

In general framework, observing and controlling on changes and unplanned scheduling in the transportation route, data connection and quality of data are challenging factors in logistics and data analysis. These reasons make also more difficult to monitoring material flow in logistics activities.

1.3.3. Logistics Data Visualization

Digital supply chain is known as a concept which means usage of digital technology for planning and fulfilling transactions in maximum level (Sanders & Swink, 2020: 44). It helps for analysing, predicting and answering alterations within logistics operations. Visualization is the significant part of this concept. As it is shown in **Figure 8**, after digitizing and connecting data, intelligent phase comes up which is one of the core attribute of digital supply chain. It is basically related with summarize and organize the big data.

Figure 8: Core Attributes of Digital Supply Chains



Source: Sanders & Swink, 2020: 45

As it is mentioned before managing and analyzing big data is quite hard and big data needed to be summarized with visualization options due to make logistics performance analysis, critiquing cost efficiency, route planning status and looking for optimization solutions. However, data is becoming larger and increasingly complex every year. Standardized data visualization is regular with developed software, for instance Excel, Phyton, R and MATLAB (Matrix Laboratory) are common, but constant and 2D images are unsuccessful to catch all over of the complicated dataset (Balzer et al., 2020: 1). That's why, usage of right tool is so important to visualize data. Visualizing data and supportive tools for it, essential for every company to ensure efficient big data management (Cremona et al., 2018: 31). When challenges are eliminated for data visualization, it helps managers to make better data interpretations and logistics decisions.

CHAPTER TWO
GENERAL INFORMATION ABAOUT COMPANY & GRAPH DATABASE

In chapter two, general information about company and department where is worked for, graph database, SWOT analysis of department and problem definition will briefly be explained.

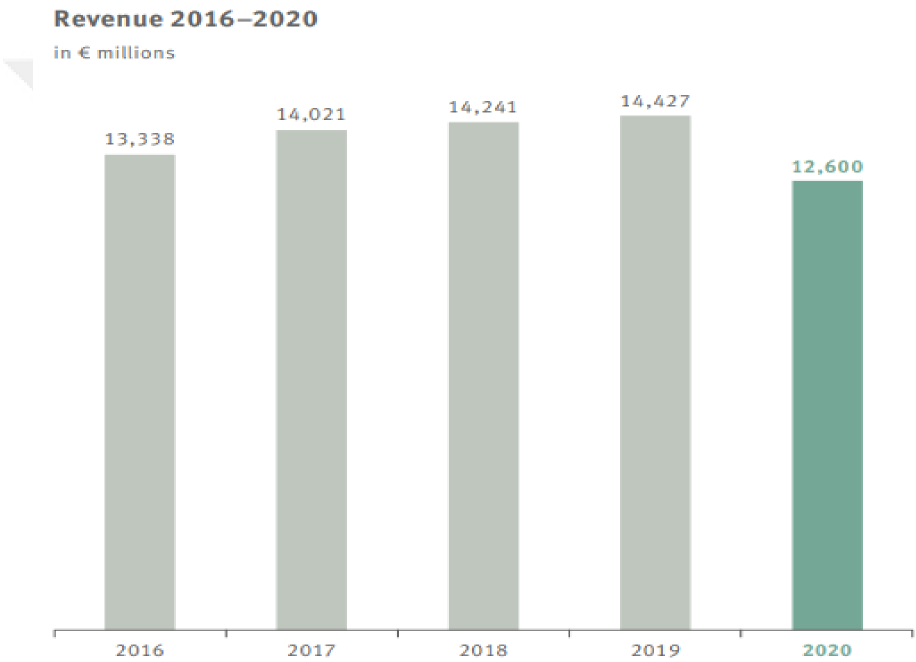
2.1. GENERAL INFORMATION ABOUT COMPANY

Schaeffler Group is a manufacturer of equipment for automotive, aerospace and industrial uses where the headquarter located in Herzogenaurach, Germany. In 1946, Schaeffler Group Industrie GmbH founded by Dr. Wilhelm Schaeffler and Dr. Georg Schaeffler. The company started to grow up after Dr. Georg Schaeffler`s invention of the needle roller bearings to contribute automotive industry in Germany in 1949. The Group became stronger with purchasing of FAG, INA and LuK and took a place as a global and industrial supplier. Today, approximately 83.000 workforce employed by Schaeffler Group and the Group is operating four different market regions where Europe, America, Asia/Pacific and Greater China. Regions is one of the significant dimensional organizational and leadership structures of the company. Other two structures are functions and divisions. Organizational structure of Schaeffler Group contains five functioning areas: “Chief Executive Officer (CEO) functions”, “Research & Development”, “Operation, Supply Chain Management & Purchasing”, “Finance & IT” and “Human Resources” (Schaeffler, Financial Statements Annual Report, 2020). On the other hand, the Schaeffler Group’s business at the helm of three divisions; Automotive Technologies, Automotive After Market and Industrial which also indicates reportable and declarable segments (Schaeffler, Financial Statements Annual Report, 2020). On following part, financial status of company is going to be examined.

2.1.1. Financial Status of Company

The group had over 14bn million-euro revenue annually in 2019. However, as it is shown in **Figure 9**, Schaeffler Group revenue decreased 12,7% in a comparison with previous year. In overall revenue growth at constant currency reduced 10,4% (Schaeffler, Annual Report, 2020). This status can be explained with decreasing demands on automotive industry due to COVID-19/SARS-CoV-2 pandemic crisis.

Figure 9: Schaeffler Group Revenue by Last Four Years

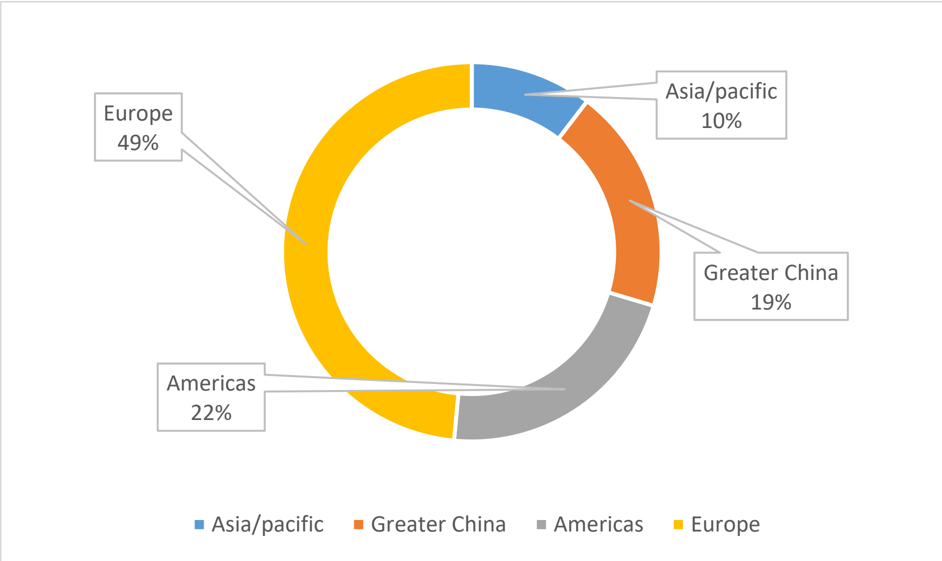


Source: Schaeffler Group Annual Report, 2020

Global world economic growth decreased 4.9% in 2020 due to the negative impact of COVID-19 pandemic crisis (IMF, 2020). Regarding to appreciations, this status in world economic growth is worse than 2008-9 financial crisis. The impact of stagnation in world economy and lockdown in global market area leaded to a decrease on Schaeffler Group’s revenue too.

Schaeffler Group is operating in four different market regions in the world and distribution of revenue by region shown as below. As it is indicated in **Figure 10**, highest revenue is ensured from Europe with 49%, America comes as secondly with 22%, 19% of total revenue is providing from Greater China and 10% is getting from Asia/Pacific as market region during 2019 (Schaeffler, Annual Report, 2019).

Figure 10: Schaeffler Group Revenue by Region



Source: Schaeffler Group Annual Report, 2019

In the following side of this study, operational areas and revenue distribution of Schaeffler Group will be explained regarding divisions.

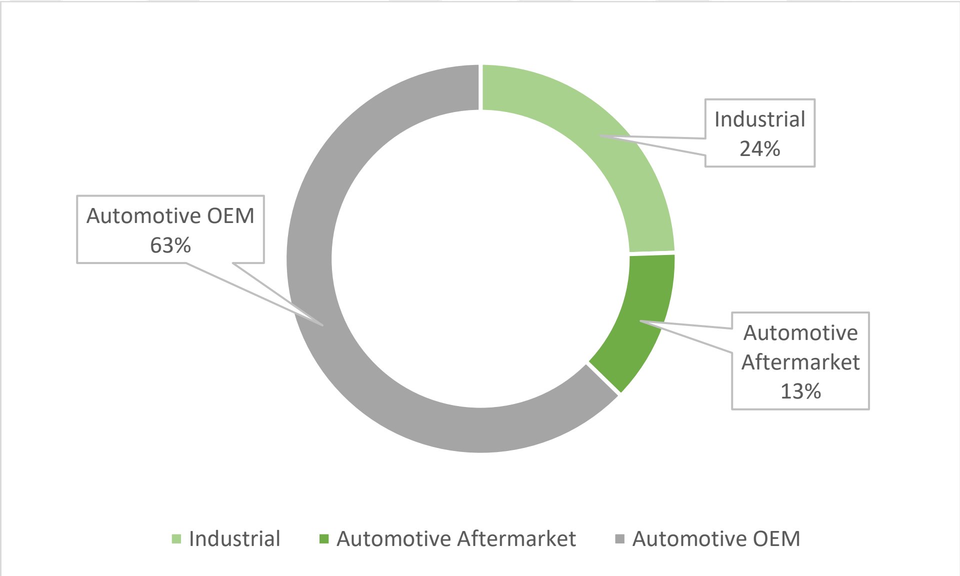
2.1.2. Operational Areas

There are three division which company is operating as it is mentioned before. The group is playing role in Automotive Original Equipment Manufacturer (OEM) area also known as Automotive Technologies, Industrial and Automotive After Market divisions. The Automotive Technologies includes transmission, internal combustion engine, chassis applications and E-mobility systems as four business divisions. The division Automotive After Market basically consists spare vehicle parts and extensive repair solutions whereas Industrial division is produces and develops components which includes rotary and linear technologies. There are eight sector segments for

Industrial division, these are aerospace, raw materials, wind, industrial automation, offroad, railway, power transmission and two wheelers (Schaeffler, Financial Statements Annual Report, 2020). To understand which division is more effective in Schaeffler Group revenue, we can examine the share of divisions within total revenue in 2019.

Figure 11 is pointing that highest share of revenue provided by division Automotive OEM (The Automotive Technologies) with 63%, in secondary level from Industrial division as much as 24% and 13% Automotive Aftermarket division in lowest level at the end of 2019 (Schaeffler, Annual Report, 2019).

Figure 11: Schaeffler Group Revenue by Division



Source: Schaeffler Group Annual Report, 2019

In 2011, Schaeffler Technologies AG & Co. KG started to play an important role as the part of Schaeffler Group. This study will consider about Schaeffler Technologies AG & Co. KG. and the data which is belongs to its department Logistics Network Design in Europe.

2.2. LOGISTICS NETWORK DESIGN DEPARTMENT

Logistics Network Design department is one of the significant departments of Schaeffler Technologies AG & Co. KG. The department is basically focusing on transportation in Europe for every transport mode. Vision, mission and strategies of Logistics Network Design department will be examined in subtitles as below.

2.2.1. Vision of Network Design Department

The Network Design Department aims to design and manage to ensure the most effective and efficient transportation network in a frame of the quality, cost and delivery flexibility. Under the light of these factors, performing in a high level in order to give answers to customers desires based on logistics activities.

2.2.2. Mission of Network Design Department

Today, Network Design department is working for lead time optimization, transport costs' reduction, just-in time deliveries, increased transparency on logistics activities in order to ensure primarily highest-level customer satisfaction and efficiency for Schaeffler Technologies AG & Co. KG.

2.2.3. Strategies of Network Design Department

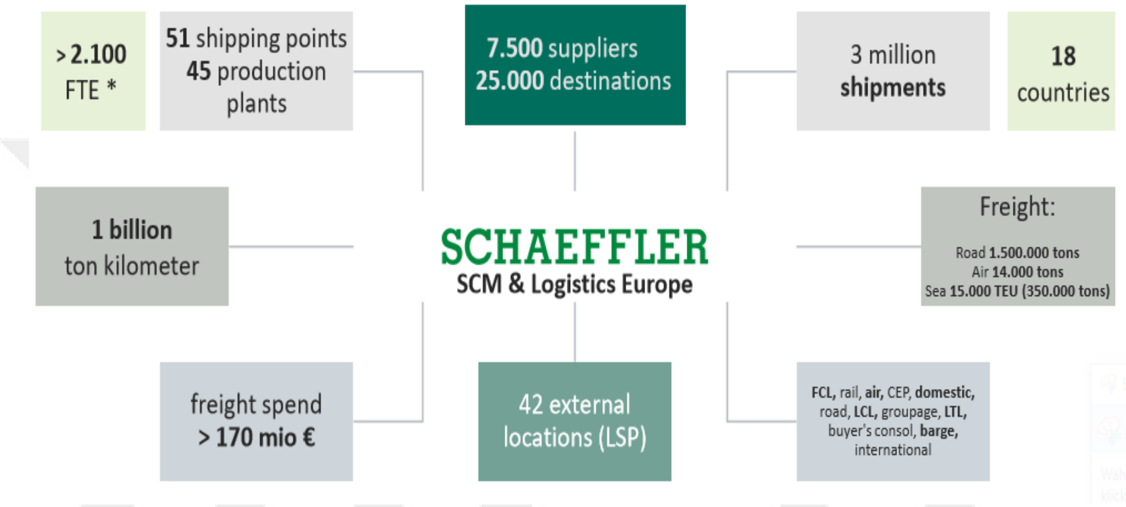
Building up new transportation concept is one of the most significant strategy of Network Design department. On this way, improving data based on delivery reliability for customers and for inter-company shipments play an important role.

Continuously optimizing and re-designing the existing transport network to provide comparable and competitive edge on lead times, transport costs, just-in time deliveries and customer satisfaction.

Moreover, to achieve transparency on all shipments, the department needs to create intelligent tools to visualize and monitor current transport performance. Thus, decision making process becomes more efficient about international forwarder.

Figure 12 is giving information and an overview about SCM & Logistics Europe regarding 2019 figures. The figure obviously shows that 3 million shipments are occurred including 18 countries in Europe with enriching 1 billion ton kilometer. 51 shipping points and 45 production plant operated for these shipments with 7.500 suppliers and more than 170 million € has been spent for transportations.

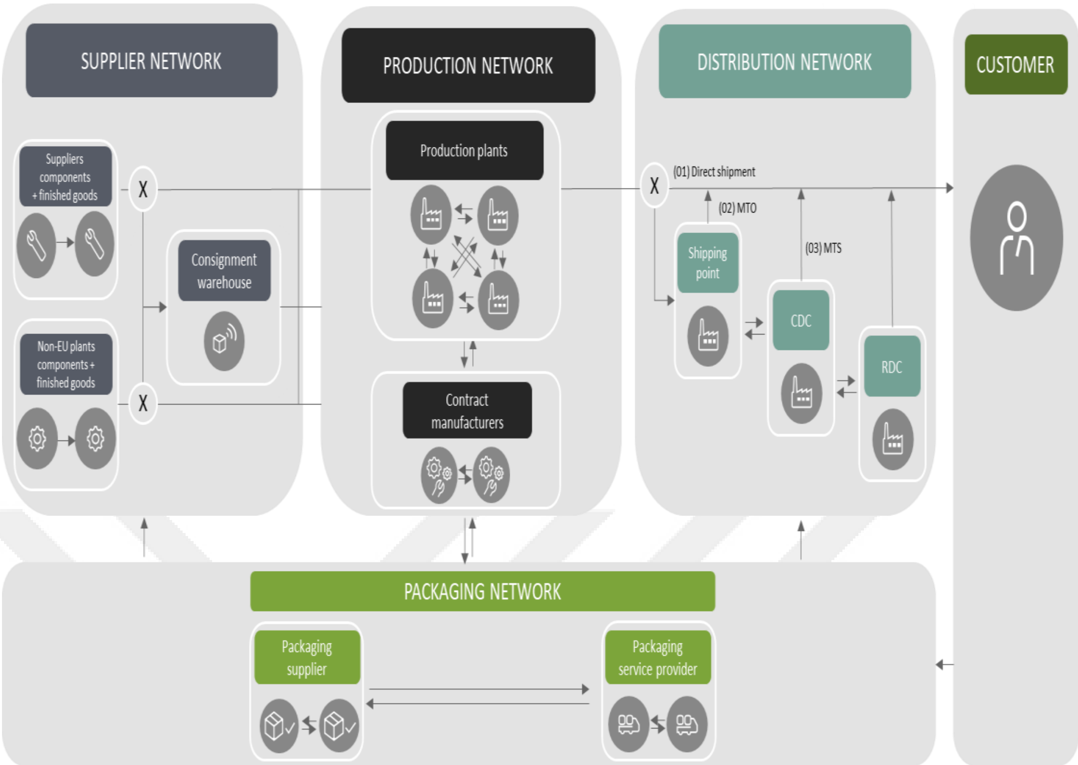
Figure 12: Schaeffler European Network Footprint 2019



Source: Schaeffler, SCM & Logistics Europe, 2019

Following figure gives better understanding for how logistics network looks like. **Figure 13** gives framework of end to end logistics network for Schaeffler Technologies AG & Co. KG. Supplier Network is start point of all logistics flow and raw materials are delivered to consignment warehouses or directly to Schaeffler production plants. After production of material, logistics in and outflow occurs with packaging suppliers and packaging service providers. When products enrich to final material level, they send to distribution centers and material flow continues until end customer.

Figure 13: End to End Logistics Network



Source: Schaeffler, SCM & Logistics Europe, 2019

Afterwards general overview of Logistics Network Design department, SWOT analysis and problem definition within the company will be inspected in the following part of this study.

2.3. SWOT ANALYSIS OF COMPANY

Senior managers of the company need to examine the firm's powers, weak points and possible options for expansion in revenue and value aspects (Mariani, 2017: 40). The benefit of SWOT analysis is a seeking for a link between exterior and interior elements to simulate new approaches (Pilar, et al., 2015: 453). Since strengths and weaknesses are basically considered on internal factors whereas opportunities and threats are examined on external causes in SWOT analysis. The internal assessment focus on all side effect within company such as workforce, capabilities, position, range of goods and services to clarify weaknesses and strengths (Dyson, 2004: 632). According to Dyson (2004), the external assessment is based on economical, governmental, social and technological impacts which are environmentally affects to the company. SWOT analysis is quite important for the mergers and acquisition strategies of the company. Therefore strengths, weaknesses, threats and opportunities of the Logistics Network Design department and its actions during the logistics activities will be examined in following part of this study.

2.3.1. Strengths

Schaeffler Technologies AG & Co. KG. is stronger part of to business-to business relations with its strong brand name in the market. As it is mentioned before, the company has huge financial flow for the logistics companies as a customer. It brings power in another level. In strategic aspect, the company has significantly located shipping points in Europe. On this way, it is possible to create centralized efficient and effective logistics activities. On the other hand, the company has workforce power with efficient and well qualified employees who are working supportively and professionally. Also, the company is operating on differentiated product range with various divisions. It helps to not rely on only one market in case decline on specified marketplace.

2.3.2. Weaknesses

Like other international companies, there are some organizational problems. For instance, organizational structure is disharmonized due to purchasing of FAG, LuK and INA in time. These problems are basically rising from functional, divisional and system interactions. Especially system interactions are quite hard because some departments are using some old self developed tools for processes. Communication flow between departments/employees is another weakness for the company. It leads to slower proceedings in workflow time to time. Decision making process is rather slow due to organizational hierarchy and size of the company. On the contrary, despite of many tools used, there is a lack of transparency on transports especially at inbound shipments. Data quality is needed to be developed.

2.3.3. Opportunities

In logistics perspective, there are many international forwarders that would like to provide services for the company, hence switching cost is not so high. Geographically, locational closeness to harbors in Europe gives consolidation opportunities in Hamburg.

Projects on harmonized team works between departments are increasing such as customs, invoice verification, logistics network design department. It gives opportunities for centralized operations and focus on common problems which are getting faced with. In technological aspect, examining usage of Robotic Process Automation (RPA) and digitalization opportunities for invoice verification department to eliminate paperwork during the process management.

2.3.4. Threats

As it is mentioned in chapter 1, political decisions are quite effective on logistics operations. BREXIT created uncertainty within logistics operations for a time. On the other hand, oil crisis is creating also negative impact on logistics activities such as in Libya occurred oil crisis in 2011, had carried its effect for a long time. Trade sanctions between countries creates also threats in global logistics sector (such as U.S. and Iran). COVID-19/SARS-CoV-2 pandemic crisis also negatively affected logistics operations globally. Especially, Chinese, American and European markets for the company.

New trends on automotive industry is pushing to changes on product sizes although it is also an opportunity for market and customer demands. In technical level, due to some deficiency in systematic operations, human mistakes are possible to happen. In operational level, updated routing implementation with another forwarder without guidelines for its specific customer creates also threats for transportations. It might lead loss of product which is rarely produced.

2.4. PROBLEM DEFINITION

Problems are defined in two stages as outbound and inbound shipments level. On the purpose of understanding of difficulties within transportations, these perspectives will be examined.

2.4.1. Outbound Shipments Level

There is lack of transparency for the optimization decision about outbound transports based on material flow. Data quality is the one of the most important reason of this situation. For instance, distance (km) information between plants and shipping points are not available like transparent flow of dispatches after one shipping point to another shipping point. On the other hand, defining real production plant is difficult time to time because there are different types of plants such as; plant for handling, plant for packaging, dummy-plant (to only book a product in the system) etc. Although each material is produced uniquely in every market region, some materials are unnecessarily shipping from one country to another country where is in another market region. It creates additional lead time and transport cost during logistics operations. On the other hand, incoterms are not clear in some points when it is EXW, Schaeffler is paying as supplier in outbound shipments. This diversification in the data creates challenges for observing material flow in logistics operations and make decisions about optimization strategies regarding requests within supply chain management by the network design department.

The products volumes are different than each other. Although one material is heavy in kilogram (kg), its` cubic meter (m^3) dimensions might be quite small. It has also a reverse situation. It leads to make decisions harder on selection of international forwarder or transport mode.

Special freight (Ausfalfracht) is another case for the company. This terminology is used in company for failed transports and dispatches for the company charged by international forwarder due to gap between planned /booked volume and real/used volume of the delivery. This situation might occur because of planning department forecasting mistakes or customers` unstable or urgent product demands.

The mentioned problem is chronic and hard to figure it out since eventuated bilateral agreement with customers.

Information flow between departments is not efficient and it leads to communicational problems. For example, dispatching mistakes rising by warehouse workers. The colleagues might forget to dispatch one product into a container, whereas its determined and included in delivery notes and related export documents. It leads huge problems for ongoing export procedures, especially on customs clearance. In order to avoid this problem, export operation team has to prepare customs clearance documents separately even these shipments dispatch in same container.

Organizational problems are also occurring from disharmonious databases. It leads to a confusion about usage of tool level of benefit and creates difficulties to work in common and integrated ground. Also, having clean data is so significant especially for freight invoice verification team but missing data on SAP about packaging standards is quite challenging.

In operational and logistics forwarder aspect, switching cost of the international forwarder is not high but deciding on new forwarder is quite risky, dealing with different forwarder currently leads to low delivery performance for customers. On the contrary, forwarder is not picking up the product on time from warehouse and being not just in time creates extremely negative impact in frame of zero-defect strategy for automotive industry.

2.4.2. Inbound Shipments Level

The lack of work and information on inbound transports. The company is mainly focusing on outbound because of expectations of the customers and customized logistics operations. We need to have transparent information on freight invoices of the inbound transports. Unlike outbound transports, detailed information is not available for inbound shipments. In the network design department, colleagues are newly preparing a project for inbound transports to have cleaner data.

Supplier differentiation leads to a complexity in Supply Chain Management. The company needs to be in contact with approximately 200 suppliers to produce one product and it is possible to clarify location of supplier unlike its dispatching location or warehouses of the products. That’s why inbound transports are more complicated than outbound shipments. To gain some solutional aspects for problems based on outbound shipments, Graph Database, graph query language; Cypher and Graph Theory will be explained in the following section of this study.

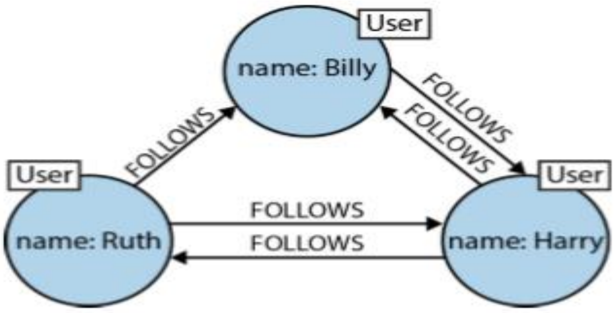
2.5. GRAPH DATABASE

In this part of study, graph database will be defined, benefits of graph database is going to clarified and Cypher which is the one of query language will be explained.

2.5.1. Definition of Graph Database

Properly, a graph gathering of corners and sides or in less dauntingly way, form of nodes and relationships which link them (Robinson, et al., 2015). In another aspect, graphs are applied to show related data which is belong to two or much more systems (Kemper, 2015). A graph database is basically a databank which uses graph patterns to portray and store data. A graph consists of nodes which are connected and organized by directional relationships. It is associated with real world and might include various scenarios because graph database fits from architecture of rocket ship to a road system, and from the supply chain to origin of the foodstuff, to health history for populaces, and beyond (Robinson, et al., 2015, S. 1). A node symbolizes entity like a customer, order, person etc. regarding the scenario of graph model as it is shown in **Figure 14**.

Figure 14: A Small Social Graph



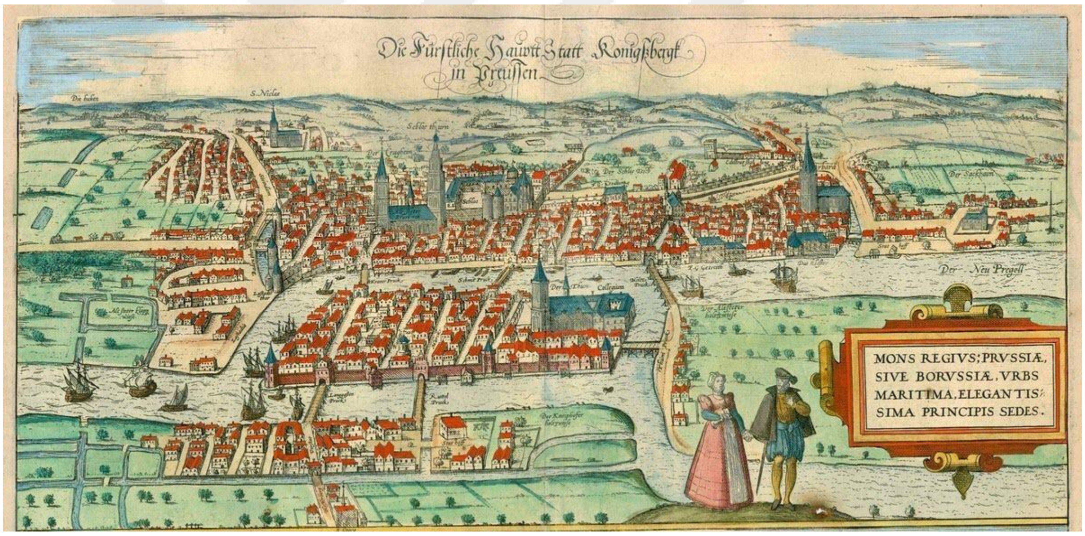
Source: Robinson, et al., 2015: 2

In following part, the graph theory and how it raised up will be examined because graph database logic is depending on this theory.

2.5.2. Graph Theory

The graph theory was raised up in a town called as Konigsberg (Prussia) city. Today the city known as Kaliningrad and it belongs to Russian Federation. The city includes two large islands due to set of Pregel River. That's why city is connected to seven bridges between two islands as it is shown in **Figure 15**. In 1736, the question of a make one city tour without passing from one bridge two times brought a problem to solve (Needham & Hodler, 2019).

Figure 15: Seven Bridges of Konigsberg

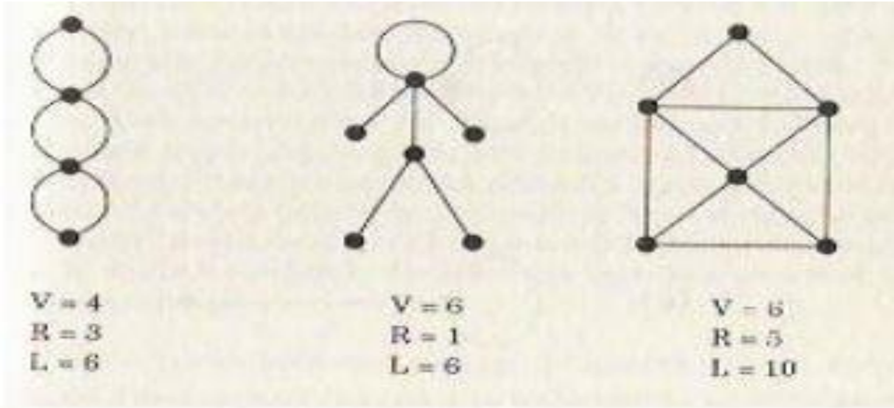


Source: Patowary, 2018

The well-known Konigsberg bridges problem solved by Leonhard Euler (1707-1783) who was a mathematician (Bellman & Cooke, 1969). Meanwhile, he developed a graph model and made an analysis to answer that why one city tour without passing from one bridge two times is not possible. His solution can be explained regarding below **Figure 16**; V shows intersections, R determines closed areas and L indicates lines on the figures. According to Euler's formula if number of lines are deducted by sum of number of intersections and number of closed areas, it always gives 1 as a result (V+R-

L=1). This formula is valid for every shape, it constitutes Graph Theory. The below structures also proof Euler’s formula. For each shapes V plus R minus L equals to 1.

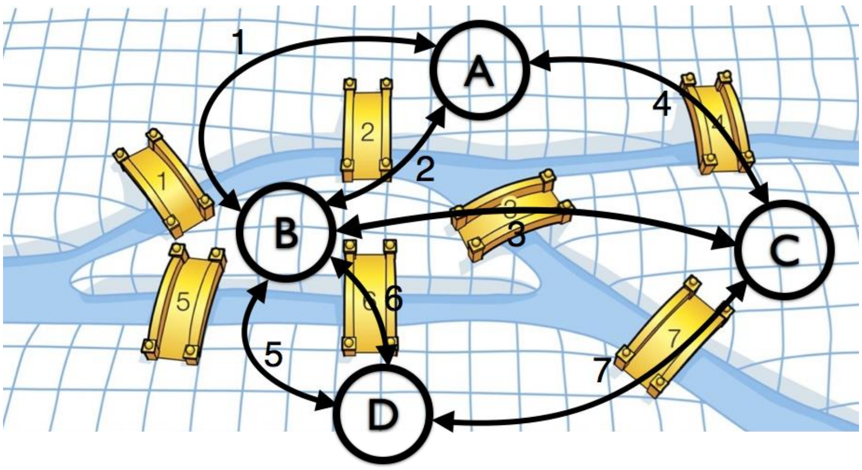
Figure 16: Euler’s Formula



Source: WALCOM, 2019

Figure 17 shows the graph version of Konigsberg. Regarding it, lands which are called *A*, *B*, *C*, *D* becomes nodes and bridges which are determined by numbers as 1,2,3,4,5,6,7 turn into relationships in graph theory and how it becomes mentioned Euler’s formula.

Figure 17: Relationship Between Euler’s Formula & Graph Theory



Source: WALCOM, 2019

2.5.3. Gains of Graph Database

A graph database performs strongly about storing and retrieving highly related data. During data retrieval performance of graph database will stay same with even millions of nodes. Manipulating the data is also flexible because there is no fixed schema between nodes and it brings an option to change, delete or create nodes, properties, and relationships. It is also possible to use different query languages in graph database (Kemper, 2015). It is also mind friendly in a comparison with other databases due to its extendable and expandable capabilities. That's why it ensures agility in database. In following part of this study, graph database will be compared with relational database and document-oriented database to understand its benefit in a better way.

2.5.3.1. Comparison of Graph Database and Relational Database

Relational databases are a selected to store large amount of data for decades due to their high performance because of atomicity, consistency, isolation and durability capacities (Holzschuher & Peinl, 2015). Relational database is more popular than graph database and it is the traditional way to store data. Traditional relational databases are generally used in intelligent transport systems (ITS) and data stored in table format and connections are ensured with primary and foreign key columns (Czerepicki, 2016). For certain type of application relational database is useful, but it is not enough with highly related data. However, graph database coherent with every types of application, especially on highly related data.

Relational database models are complicated. Querying one entity is easy in relational data model, but when every detail was needed for entity, queries becomes long and performance of retrieving data gets slow. That's why relations are not carrying high importance in relational database. Graph database has flexible schema, but it is not same for relational database Also, queries and structure are more understandable in graph database. Relational database fetches the data with joins whereas graph database fetches with a pattern (Robinson, et al., 2015). **Table 1** also shows differences between relational database and graph database.

Table 1: Relational Databases Versus Graph Databases

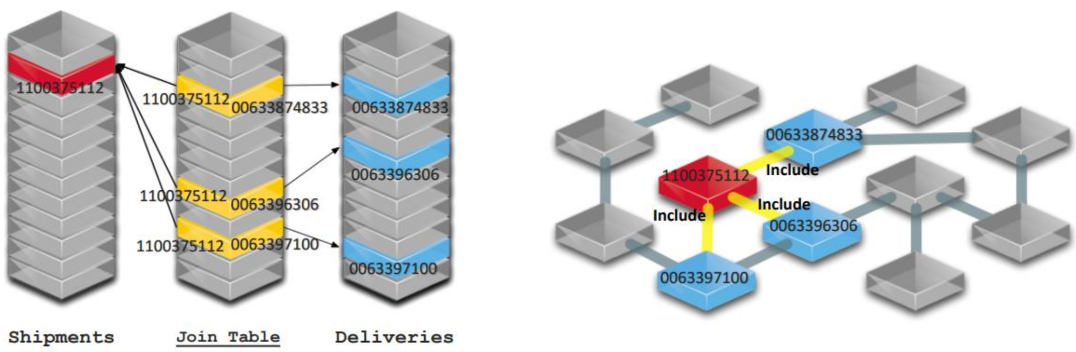
Relational Database	**Graph Database**
Use Tables to store data	Stores Data as Nodes
Schema with nullable	No schema: no need to work with null value
Relations with foreign keys	Relation is first class citizen (no duplicated data)
Related data fetched with joins	Related data fetched with a pattern

Source: Robinson, et al, 2015

Advantage of relational data base is being capable to use on uniformly structured data, it makes databases very suitable for reporting and calculations in one table. However, disadvantage of it can be summarized being complicated, costly and taking longer time on retrieving data in a comparison with graph database.

Figure 18 shows an example from Logistics Network Design department and indicates how it is visible to see data in graph database when it is compared with relational database. Shipment number which is clarified with red color includes several delivery numbers where is shown in blue color. In relational database one join table (defined with yellow color) needs to be created to connect two different records whereas only relationships need to connect in graph database.

Figure 18: An Example to Compare Relational Database & Graph Database



Source: Schaeffler, SCM & Logistics Europe, 2019

2.5.3.2. Comparison of Graph Database and Document Database

The document-oriented database model is depending on key paradigm. The data capsulizes, encrypts and stored in a semi structured format like XML and JSON (Bouaziz, et al., 2019). Document database is a storage for documents all related data in one entity. It is good when all related data stored in one document. However, duplication of data is not something which is avoidable. Since, master data item copied to document and it leads duplicated data. **Table 2** indicates differences between document database and graph database in detail. The main source of the graph database is nodes whereas document-oriented database is document. Both databases have no schema which need to be considered on null value. In document database, relationships are designed with foreign keys, meanwhile relationships are first class citizens in graph database. On the other hand, document database fetches the data with join tables, but graph database follows a pattern.

Table 2: Document Databases Versus Graph Databases

Document Database	**Graph Database**
Document	Nodes
No schema	No schema
Relations with foreign keys or embedded	Relations is first class citizen
Related data fetched with joins or embedded	Related data fetched with a pattern

Source: Robinson, et al., 2015

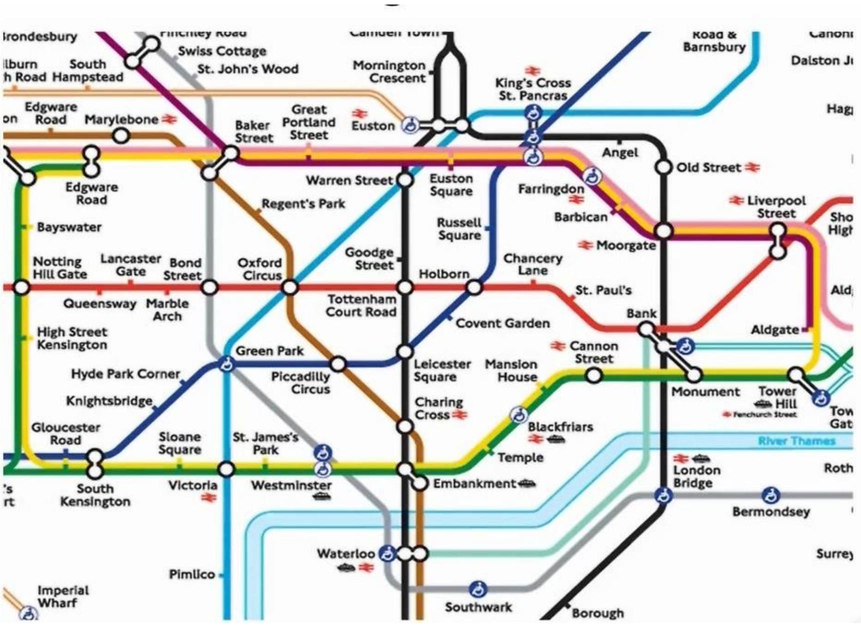
Due to all these comparisons with other databases, graph database is selected in this study to build a tool.

2.5.4. Graph Database in Logistics

Several tools need to be used to assure making effective and efficient decisions for resiliency to potential risks and responding transport operations in logistics management (Sobczak, et al., 2018). On database systems, demands are increased and it caused a raised prevalence in niche fields like networks of graph, social networking or logistics structure based on usage of data systems (Wallden & Ozkan, 2016).

Figure 19 presents subway stations in London as a used case in Graph Database. In this example, stations show nodes, directions indicate relationships, and these relations ships includes some attributes like distance and arriving and leaving time information. Regarding this instance, it is possible to calculate shortest route with the help of graph database. It is basically a sign of how graph database can be used in logistics sector.

Figure 19: Graph Database Used Case: Subway Stations in London



Source: Depeau, 2019

2.5.5. Cypher Query in Graph Database

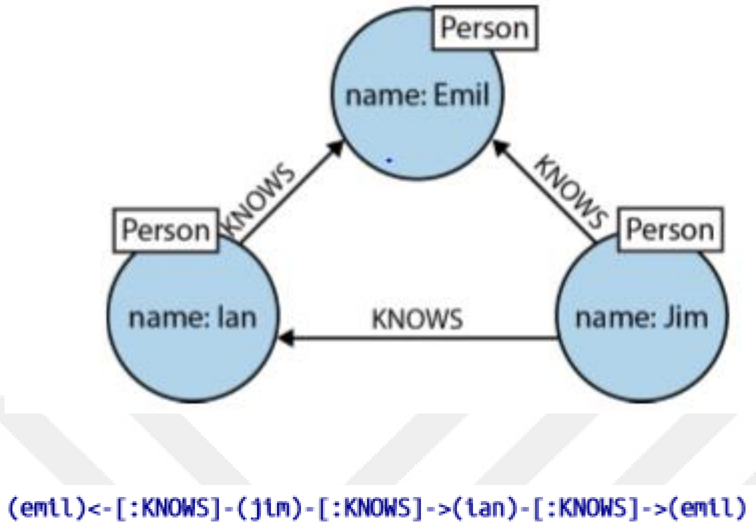
Cypher is the graph query language which helps express to data querying. The below **Figure 20** is a basic model for three persons who knows each other. It is so obvious to see the data with their label as “name” and relationships “knows” directions. In the below part, how to create and delete nodes and relationships, filter to data, add extra information into data and how can we modify the data which already existing. According to Kemper (2015), basic syntax in Cypher query looks like below;

( ) Node

{ } Properties

[ ] Relationships

Figure 20: Graph Database Example: Nodes & Relationships



Source: Robinson, et al., 2015: 28

2.5.5.1. Creating Nodes and Relationships

To create an n node simply, below query is used. The Return clause is necessary to get an outcome in case that node is created earlier (Panzarino, 2014). Since, there might me another n node which is created with a different id number. It is possible to add a label to node n when it is created. So, in our below query shows that n node labeled as "Person".

```
CREATE (n)                    CREATE (n:Person)
RETURN n                      RETURN n
```

There might be also multiple labels in one node. Thus, we have a better opportunity include more information into our model and filter regarding our analysis. The below query means that n node is "person" at the same time "inactive".

```
CREATE (n:Person:Inactive)
RETURN n
```

Above nodes which are created are not including any properties. At the beginning, our overview figure included “name” property but how can we create it? As below query, person node has properties with its name and age information.

CREATE (n:Person:{name:”Jim”, age:”28”})
RETURN n

2.5.5.2. Modifying Existing Data

To modify and get existing data, match, merge and set clauses are possible to use in Cypher query language.

Match clause ensure to get information about nodes, relationships and properties with carrying filter function in the data set (Kemper, 2015: 40). For instance, below query filters a “person” node with property “name” Chris and get with only this node with return clause.

MATCH (n:Person{name: “Chris”})
RETURN n

It is also possible to filter indicated relationship between two nodes as below query node a is related to node b;

MATCH (a)- [r: RELATED]->(b)
RETURN a, r, b

Meanwhile, match clause is providing filter option in the data, merge clause ensures additional function. If there is no information for defined query, merge clause creates missing information and filters data at the same time. That’s why it should be used carefully because it creates opportunity to manipulate data. Merge clause is mix version of create and match clauses (Kemper, 2015: 51). For example, in current data node “bill” available but there is no information about property “age”. Below merge clause catch the node “bill “and creates “age” property for the node “bill” at the same time.

```
MERGE (bill {name: “Bill”, age: 28})
RETURN bill
```

On the other hand, set clause is also used commonly in Cypher query. It ensures to update labels on nodes, relationships and properties. However, it can not be used without match clause because data should be filtered first to catch and make desired updates on labels. Below Cypher query would update node “n” with new “username” property when property is already existing unless it creates “username” property and add into node.

```
MATCH (n {name: “Chris”})
SET n.username = “chrisdkemper”
RETURN n
```

2.5.5.3. Deleting Nodes and Relationships

Delete clause helps to remove nodes and relationships. It is a significant Cypher clause as much as match, merge and set clauses. Because there might be wrong or unnecessary information within data. Firstly, match clause should be used to catch node which is intended to delete (Kemper, 2015: 43). Below query indicates how to delete node “n”;

```
MATCH (n: RemoveMe)
DELETE n
```

However, “n” node can have a relationships and above query does not help on this case because it can not delete “n” node unless there is no delete clause for relationship as well. On that way node “n” remains. To delete node which has also relationship, below query can be used and on this way node “n” could be deleted.

```
MATCH (n: RemoveMe)-[r]-( )
DELETE n, r
```

There are many Cypher clauses to use in case of the structure of the data and which kind of analysis is would it wanted to make. However, main clauses are explained in this study to give ideally how to write cypher queries. It was shown that how Cypher query is readable and easy to write in graph database. In the following part of the study, new chapter will include data analysis and results and they will be explained in detail.

CHAPTER THREE
ANALYSIS AND RESULTS

The third chapter contains objective of the research, related research questions, methodology, data analysis and results.

3.1. OBJECTIVE OF THE RESEARCH

This project aims increased transparency, visualize worldwide outbound transports, better understanding material flow and performance quality of international forwarders based on graph database. Methodology is the data modelling in relational database from SAP and data migration to graph database in order to ensure pellucidity on outbound shipments. Our outcome will be visualized material flow and regarding our consequences, better optimization solutions will be raised up for colleagues to focus on. As an addition to main objective of this study, there are some sub-objectives; In current status, reporting opportunities are limited in material, delivery and shipment level for the colleagues. There is a need for occurring a combined platform for three of them. Also, this platform will include Key Performance Indicator (KPI) reports for international forwarders.

3.2. RESEARCH QUESTIONS

The research questions are constructed as below:

R_1: Is it possible to visualize material flow in outbound transports?

R_2: How can we have combined platform in material, delivery and shipment level?

R_3: Is it possible to find alternative plant for the same material which is produced for the same customer?

R_4: Is it feasible to find shortest path between plants and end customers?

These questions will be tried to explain in a practical way. In the bellow methodology part of this study, sampling and data modelling are going to clarify briefly.

3.3. METHODOLOGY

In the following part, methodology of this study will be explained comprehensively with sampling and modeling of data and data migration sections.

3.3.1. Sampling

Sampling is started with using small portion of data which is including shipments which are sent to only Turkey for 3 months. This sampling aimed to understand possibilities of uploading data via CSV file. On this concern, the performance of the graph database is checked without transferring the data from SQL Server directly. After small testing phase, huge and focused data is determined. Concentrated data is based on final material which is transported as outbound shipment for one-year data of 2020. There are three main push factors for why this portion of data is selected. To have actual and latest data in case of eliminating some shut down shipping points, outbound data is cleaner than current inbound data in availability level and final material to focus on customer satisfaction based on logistics activities.

3.3.2. Data Modeling

Data is modelled with an Entity Relationship Diagram (ERD) and collected by using SAP(EP1) and SQL Server. To build the architecture of data model; SAP SE16 data browser used to bring table for each node. Each table connected to each other with joint tables and foreign keys in Relational Database. **Table 3** shows SAP transactions which are used to create below nodes. There are 13 nodes in the data model, other three nodes created via using internal data source in the department. The data is quite comprehensive in the level of variety, since project will be continued for a time in the Network Design department.

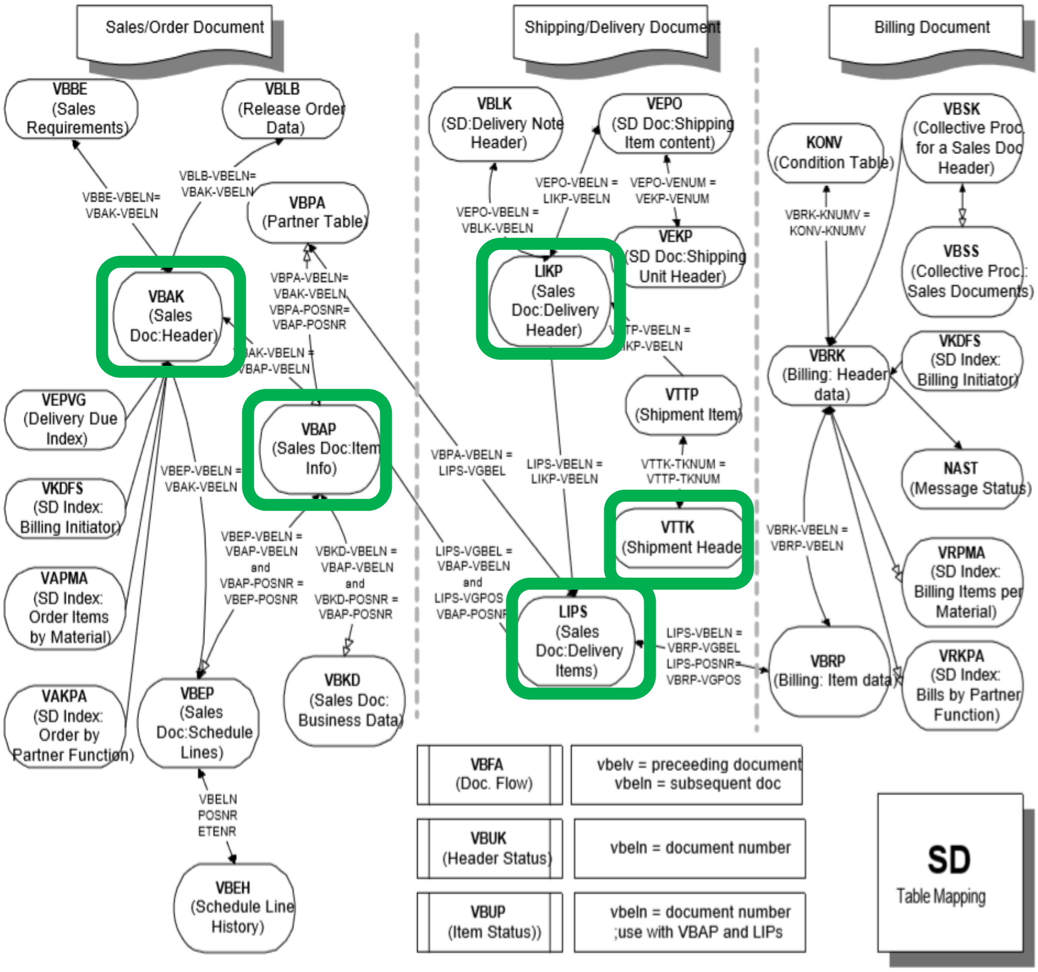
Table 3: SAP Descriptions Used for Nodes

Name of Node	SAP Table Name
1.Product	LIPS, MARA
2.Delivery	LIKP
3.Shipment	VTTK
4.Sold_To (AG)	KNA1, KNVP
5.Ship_To (WE)	KNA1, KNVP
6.Plant (PWERK)	T001W
7.Shipping Point (ShipPoint)	TVST, TVSTT
8.Forwarder	LFA1
9.Address	ADRC
10.Place	ADRC

Source: SAP Tables, 2021

Figure 21 shows a well-known SAP SD table mapping which is belong to sales and distribution connections. In SAP R/3, we do not have information in logistics level, but it is possible to get data from SAP S4 HANA. In this study, determined mapping table logic applied to data architecture of logistics flow. The order created by user in VBAK (sales document, header data) as shown in below figure. Under the VBAK, gives data about address information (ADRC) as a sales document item data. When the order created, delivery information created in LIKP (delivery table, header data) with delivery positions and it will include also material data in LIPS (delivery table, item data). LIPS delivery position insists level of product as well such as final material, packaging material, handling unit information etc. When delivery created for assigning who will ship it and shipment number will be created in warehouse system, candidate freight forwarder based on two different tables in SAP and will be assigned to route in this tables will be connected to VTTK (shipment table, header data). So below figure shows basically, which SAP description are used for data modelling mainly.

Figure 21: SAP: Sales and Distribution



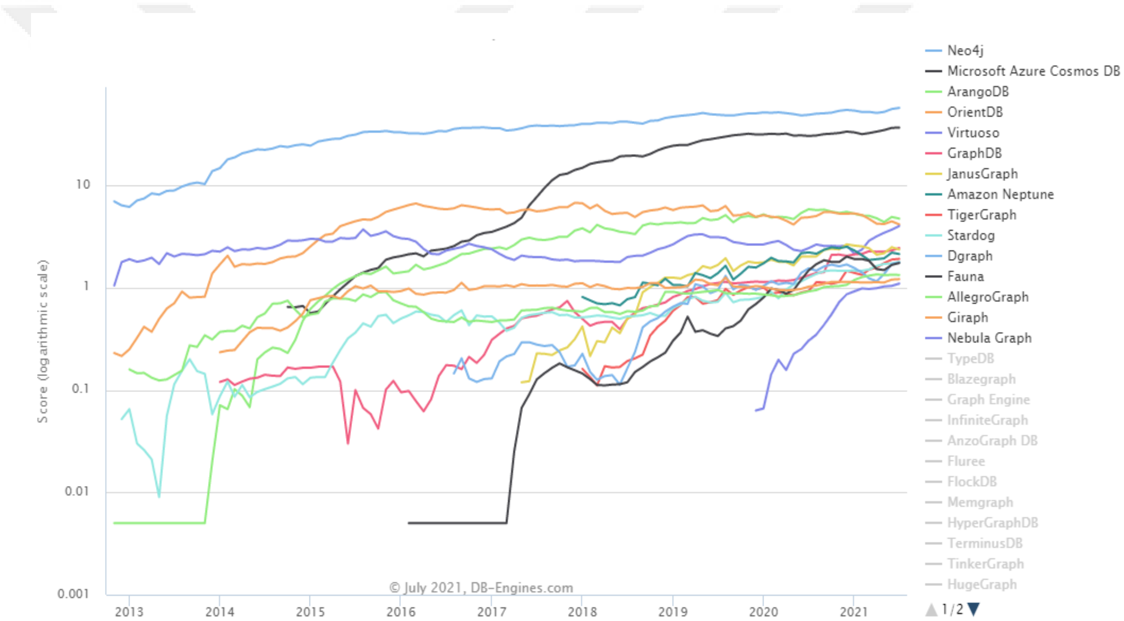
Source: Solomon, SAP Table Relations version 1.0.0

3.3.3. Neo4j and Data Migration

Database management system (DBMS) is used to organize data in traditional database and connect with selected database. In the wide market, it is possible to use many DBMSs like MySQL, CouchDB, Microsoft SQL Server, PostgreSQL or else Neo4j (Kemper, 2015). Neo4j started as an open-source project in 2003 by Neo Technology. In time, it provided an open-source graph database which is implemented in Java.

The platform has significant pillars like performance, high availability, agility, security and developing efficiently meanwhile retrieving data (Robinson, et al., 2015). On the other hand, Neo4j ensures many functions such as mathematical aggregate functions, collect to create lists, data formatting capabilities, logic constructs, list predicate logical functions and query profiling in graph database (Kemper, 2015). Therefore, it is highly used as graph database management system (GDBMS). **Figure 22** indicates that Neo4j has is highest rank level of usage in comparison with other GDBMS since 2013.

Figure 22: Database Engines Ranking of Graph Database Management System



Source: DB-Engines Ranking, 2021

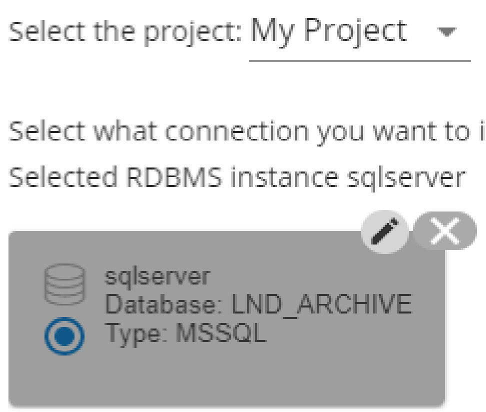
Many companies are using Neo4j because there is a huge growing adaption in graph database. It is possible to use graph database in various business sectors for big data management. Since, graph database is giving opportunity to analyze data with visualization option and better understanding rather than relational database and table format. IBM, AstraZeneca, Caterpillar, Airbus, U.S. Army are some customers of Neo4j (Neo4j Customers, 2020).

As it is mentioned in previous chapter, Cypher query is used in Neo4j. Cypher is intended to use and read easily. It is defined on Neo4j website as like SQL on declarative and query language textuary way (Kemper, 2015: 15). It makes easy to write and read queries descriptively in a comparison with other graph database platforms' query languages.

There are some reasons of necessity to use Neo4j in this study. The current platform which is used by Logistics Network Design department gives reports in separated level for the data on all transportations. These levels are material, delivery and shipment and they are not integrated to each other. For instance, cost simulation is only working on shipment level, there is no opportunity make cost analysis in material level automatically. Also, this tool is working regarding to excel file and it makes data limited. Since, excel can only report one million data in maximum level. On the other hand, we need to analyze much more amount of data to find out better optimization solutions. That's why neo4j will be more beneficial for our colleagues.

To ensure data analysis firstly, data migration was needed. For the data migration, Extract, Transform, Load (ETL) tool helped to transfer the data from SQL server. In Neo4j, ETL tool ensured to create nodes and relationships names and saving to mapped data. Online direct import from SQL to graph data base is ensured with this tool. **Figure 23** shows how to connect Neo4j to SQL server to bring intended data portion.

Figure 23: ETL Tool "Load Data Model"



Source: Neo4j Desktop-1.2.8-ETL Tool, 2020

As it is shown in **Figure 24**, it is possible to map our meta data via ETL Tool to create relationships between determined nodes.

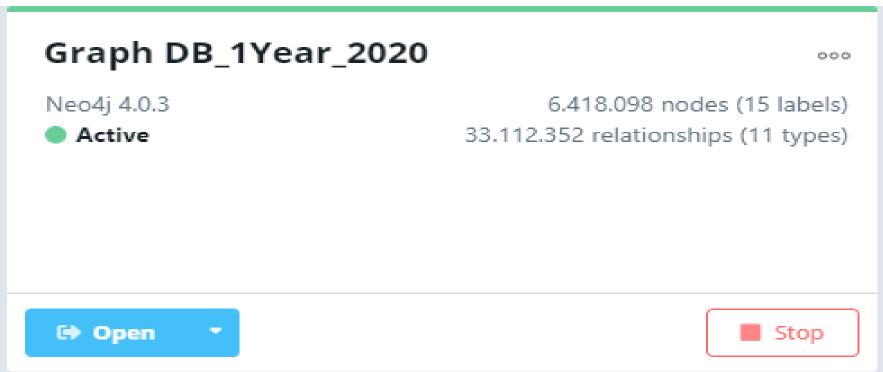
Figure 24: ETL Tool “Explore and Change Metadata”

Source: Neo4j Desktop-1.2.8-ETL Tool, 2020

After all process, we can import our data into Neo4j. It is possible to create graph database via remote connection either desktop version. If these ways are not proper for the user, it is also fine that uploading CSV files with a small portion of data. However, in this study, creating desktop version of graph data base is selected due to large volume of the data.

As below, **Figure 25** shows that in final Graph Database which includes more than 6 million nodes and moreover 33 million relationships between these nodes. Nodes had different 15 labels and there were 11 different types of relationships.

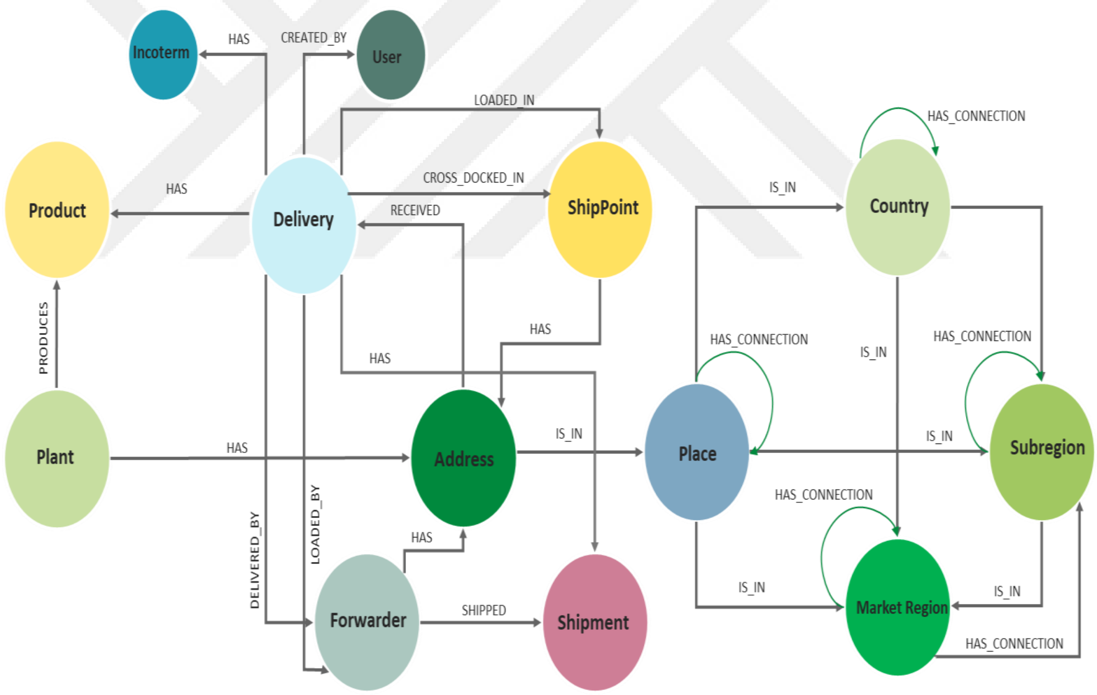
Figure 25: Graph Database: Number of Nodes and Relationships



Source: Neo4j Desktop-1.2.8, 2020

In this study, data model is created as below. To get schema for the mapped data “*call apoc.meta.graph*” is used as query. Below data model shows nodes and relationship in logistics data. On this case tables became labels, columns in the table turned to properties, records in tables indicate nodes and foreign keys which are ensuring connection between two tables in relational database come to be relationships in graph database. **Figure 26** demonstrate 13 nodes and 26 relationships between nodes. Some relationships named in same way to decrease complexity during the writing Cypher query. In overall, delivery created by which user, delivery has which incoterms, delivery has which product, delivery loaded and delivered by which forwarder, delivery loaded and cross docked in which shipping point, plant produces which product, forwarder shipped which shipment, address received which delivery; plant, forwarder and shipping point has which address, these addresses is in which place, these places is in which country, subregion and market region.

Figure 26: Created Schema of Logistics Data in Schaeffler



Source: Neo4j Desktop-1.2.8, 2020

In the following part of this study, nodes and relationships will be explained in a detail that which kind of information they include. In above shown data model is in graph model level however there are many properties and attributes are hidden in this data model.

3.3.4. Nodes and Relationships

The data model of the study involves several nodes and properties. In below side, properties name of the nodes will be shown with their ordinal positions. Ordinal positions mean how the properties sequenced with their name in graph database. To demonstrate briefly how the data looks like in graph database, the company`s current data archive is used to bring in table version. Since, it is hard to show hidden properties which are included in graph database on the paper.

3.3.4.1. Plant Node and Relationships

Plant node is prepared by using T001W SAP transaction. As it is shown in **Table 4**, plant node includes for every plant its id, code, name, address, second name and plant type. Plant might act like a customer, also there are five types of plant in the company system. All of them does not act like a real producer. Producer plant indicated with P. Other types are Landegesellschaft (L), Handelhaus (H), Packmittel-Werk, Dummy-Werk(D). All of them indicated in German language. In order to ensure better understanding, L is landing location, H shows handling and R signifies packaging unit. On the other hand, D is used in order to book some products in the system, they are not real producer but creates complication for the data. That’s why we had to determine these plants with properties as named plant type in node level.

Table 4: The Properties of Plant Node

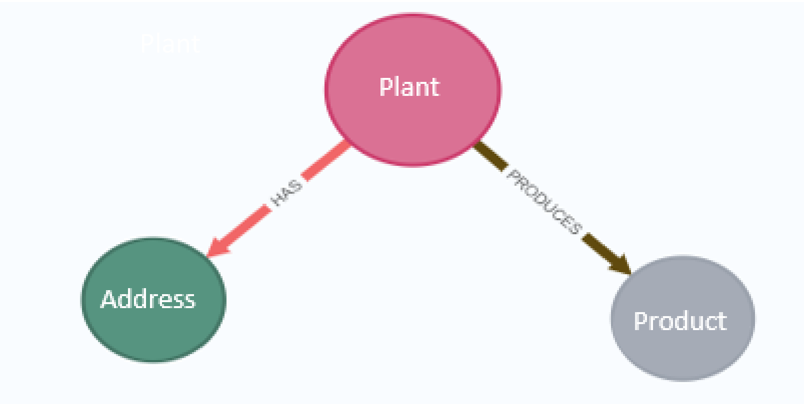
COLUMN_NAME	ORDINAL_POSITION
ID	1
Code	2
Name	3
Address	4
Name2	5
Plant_type	6

Source: Schaeffler Data Archive, 2020

Figure 27 shows as starting plant node linked to product and address nodes via “PRODUCES” and “HAS” relationships. “PRODUCES” relationship indicates plant produces which products. “HAS” relationships helps to get data about address of plants. It is possible to see relationships where product nodes are origin via below query.

```
match(a:Address)<-[r1:HAS]-(p:Plant)-[r:PRODUCES]->(pr:Product)
return p,r,pr,a,r1 limit 1
```

Figure 27: Relationships of Plant Node as Origin



Source: Neo4j Desktop-1.2.8, 2020

3.3.4.2. Product Node and Relationships

Product node is prepared by using of SAP Table LIPS. In order to get material number MARA is selected to bring data. Basically, final material is used to decrease data complexity and focus on data in a level of customer satisfaction. ERZE, is called in the company as final material and it is the final version of a product to deliver end customer. Table`s column names for Product Nodes are id, code, name, gross weight, net weight, unit weight, product hierarchy, division (shows that product is industrial, automotive or automotive aftermarket) information as it is shown in **Table 5**.

Table 5: The Properties of Product Node

COLUMN_NAME	ORDINAL_POSITION
ID	1
Code	2
Name	3
Gross_weight	4
Net_weight	5
Unit_weight	6
Product_hierarchy	7
Division	8

Source: Schaeffler Data Archive, 2020

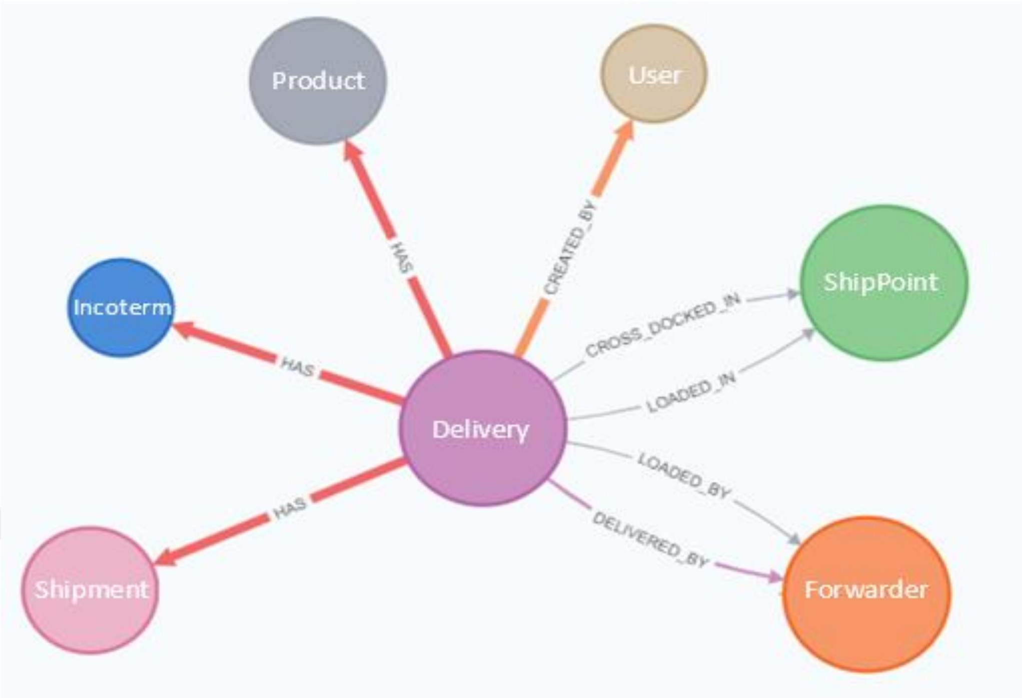
Product node does not have any relationship where it is a start node.

3.3.4.3. Delivery Node and Relationships

LIKP table is used as SAP ABAP table field to generate Delivery node. This node includes 25 properties because delivery nodes plays important role during the transportation with those properties. For the continuity of the project in future, properties variety is quite high in the data model. However, the main properties are used in this study. In **Figure 28**, there are eight relationships where delivery node is origin regarding direction of link. Also, below query helps to get these relationships with other nodes of delivery node.

```
match(n:Delivery)-[r:CROSS_DOCKED_IN]->(sp:ShipPoint)
match(n)-[r1:LOADED_IN]->(sp)
match(n)-[r2:DELIVERED_BY]->(f:Forwarder)
match(n)-[r3:LOADED_BY]->(f)
match(n)-[r4:HAS]->(p:Product)
match(n)-[r5:HAS]->(s:Shipment)
match(n)-[r6:HAS]->(i:Incoterm)
match(n)-[r7:CREATED_BY]-(u:User)
return n,sp,f,u,p,s,i,r,r1,r2,r3,r4,r5,r6,r7 limit 1
```

Figure 28: Relationships of Delivery Node as Origin



Source: Neo4j Desktop-1.2.8, 2020

On **Table 6**, it is possible to see properties of delivery node. Code indicates delivery number, freight payer shows which party takes responsibility of transportation cost, harmonized weight expresses delivery weight in kg and tons, delivery value in eur gives information about sales value of indicated shipment, shipping condition includes details about making standard, express, or special freight, handling unit quantity shows number of pallets in delivery, ship to column is indicating delivery sent to physically which receiver whereas sold to shows number and name of purchaser of mentioned delivery.

Table 6: The Properties of Delivery Node

COLUMN_NAME	ORDINAL_POSITION
ID	1
code	2
Delivery_manual_batch	3
Order_combination_LIKP	4
Sales_organization	5
Freight_payer	6
Special_processing_#	7
Proportion_on_shipmen_quantity	8
HU_quantity	9
OrderLines	10
Delivery_product_class	11
Delivery_combination_ZLKDSCH	12
Dispatch_mode	13
Harmonized_weight_KG	14
Harmonized_weight_TON	15
LD_Delta_delivery_created_planned	16
Delivery_combination_LIKP	17
CD_flag	18
Delivery_value_EUR	19
Shipping_condition	20
Ship_to_#	21
Ship_to_name	22
Sold_to_#	23
Sold_to_name	24
Shipping_condition_desc	25

Source: Schaeffler Data Archive, 2020

Table 7 shows attributes of CREATED_BY relationship between Delivery & User Nodes. It includes information about user ID, delivery ID and date of creation of related delivery number as an order.

Table 7: The Attributes of CREATED_BY Relationship

COLUMN_NAME	ORDINAL_POSITION
ID	1
User_ID	2
Delivery_ID	3
Created_Date	4

Source: Schaeffler Data Archive, 2020

Another relationship is CROSS_DOCKED_IN where delivery node is origin. This relationship is between delivery and shipping point nodes. It included in this project for future possible analysis. On **Table 8**, shipping point data is included for cross docked deliveries to catch crossdocking is occurred based on which shipping points.

Table 8: The Attributes of CROSS_DOCKED_IN Relationship

COLUMN_NAME	ORDINAL_POSITION
ID	1
Delivery	2
ShippingPoint	3

Source: Schaeffler Data Archive, 2020

For deliveries, it should be clear that deliveries loaded on which planned and actual day and which shipping point. For that reason, LOADED_IN relationship between delivery and shipping point nodes includes mentioned information as below on **Table 9**.

Table 9: The Attributes of LOADED_IN Relationship

COLUMN_NAME	ORDINAL_POSITION
ID	1
Delivery	2
ShippingPoint	3
LD_A_Loading_Date	4
LD_P_Loading_Date	5

Source: Schaeffler Data Archive, 2020

In one level delivery node link to forwarder node with DELIVERED_BY relationship. It includes delivery actual delivery date and delivered by which forwarder information as it shown on **Table 10**.

Table 10: The Attributes of DELIVERED_BY Relationship

COLUMN_NAME	ORDINAL_POSITION
ID	1
Delivery	2
Forwarder	3
LD_A_Delivery_Date	4

Source: Schaeffler Data Archive, 2020

In another level delivery node link to forwarder node with LOADED_BY relationship. Forwarder can differentiate in case of delivering to one place and loading in same place for the shipments. That's why, there is two different relationships between delivery and forwarder node in same directions. **Table 11** shows attributes of LOADED_BY relationships between delivery and forwarder nodes with insisting actual loading date regarding determined delivery and forwarder.

Table 11: The Attributes of LOADED_BY Relationship

COLUMN_NAME	ORDINAL_POSITION
ID	1
Delivery	2
Forwarder	3
LD_A_Loading_Date	4

Source: Schaeffler Data Archive, 2020

In the data modelling, we have more than one relationship which named as HAS. This way is selected to decrease complexity in data structure. One of HAS relationship is between delivery and incoterm nodes. On **Table 12**, it is possible to see type of incoterm and its agreed location to clarify transport cost and risk belongs to which side; shipper or receiver.

Table 12: The Attributes of HAS Relationship: Delivery & Incoterm Nodes

COLUMN_NAME	ORDINAL_POSITION
ID	1
Incoterm	2
Delivery	3
Incoterm_Place	4

Source: Schaeffler Data Archive, 2020

On **Table 13**, another HAS relationship is shown between delivery and product node. It is important to catch material flow during logistics activities because it includes information about indicated delivery includes which materials (products) with its unit weight, net weight, gross weight, delivered quantity and plant information which produced those materials.

Table 13: The Attributes of HAS Relationship: Delivery & Product Nodes

COLUMN_NAME	ORDINAL_POSITION
ID	1
Delivery	2
Product	3
Plant	4
Delivered_Quantity	5
Gross_weight	6
Net_weight	7
Unit_weight	8

Source: Schaeffler Data Archive, 2020

Below **Table 14** contains attributes of HAS relationship between delivery and shipment nodes. Since, one shipment number might include several different delivery numbers. That's why, this relationship is created in the data model.

Table 14: The Attributes of HAS Relationship: Delivery & Shipment Nodes

COLUMN_NAME	ORDINAL_POSITION
ID	1
Delivery	2
Shipment	3

Source: Schaeffler Data Archive, 2020

3.3.4.4. Shipment Node and Relationships

In the architecture of graph data model, shipment and delivery nodes are different than each other because in company on going system, every shipment numbers contain various delivery numbers as it mentioned before. That's why we need to have two different nodes for delivery and shipment information. This node includes basically cost information which is invoiced of logistics operation. On the other hand, distance (km), route, transport mode, type of transport (express or standard shipment) and volume information of dedicated shipment. It is possible to see properties of shipment node regarding ordinal position on **Table 15**.

Table 15: The Properties of Shipment Node

COLUMN_NAME	ORDINAL_POSITION
ID	1
Code	2
Type_of_transport	3
Transport_mode	4
Route	5
LC_inv_costs	6
LC_inv_from	7
LC_inv_service	8
LC_inv_mode	9
LC_inv_consol	10
LC_cal_costs	11
LC_costs	12
LC_source	13
LC_per_KM	14
LC_per_Sales	15
LC_per_KG	16
LC_inv_volume	17
LC_inv_loaded_meters	18
LC_inv_distance	19
LC_inv_volumeweight	20
LC_inv_weight	21
LC_inv_verification_status	22

Source: Schaeffler Data Archive, 2020

Shipment node does not have any relationship where it acts as origin node.

3.3.4.5. Shipping Point Node and Relationships

Shipping point node has below properties which is revealed on **Table 16**. Code, name and address are significant information for shipping point. There are numerous shipping points where Schaeffler uses in logistics activities. Especially address information would be helpful for optimization solutions.

Table 16: The Properties of Shipping Point Node

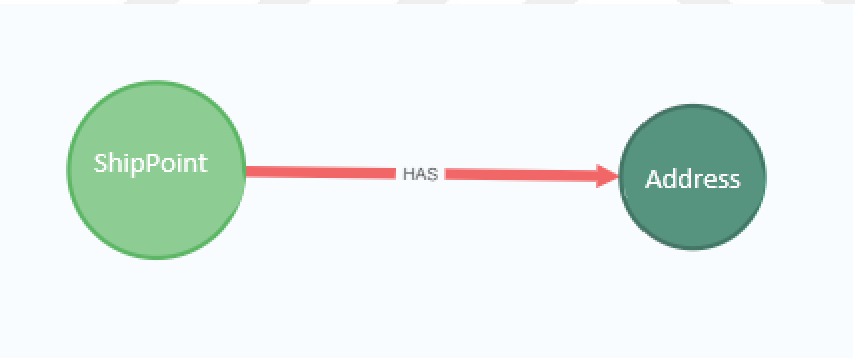
COLUMN_NAME	ORDINAL_POSITION
ID	1
Code	2
Name	3
Address	4

Source: Schaeffler Data Archive, 2020

In data model, there is only one relationship where shipping point is origin node. There are not any attributes in this relationship. It is possible to get HAS relationships with below query and the relationship looks like below in **Figure 29**.

```
match(sp:ShipPoint)-[h:HAS]->(a:Address)
return sp,h,a limit 1
```

Figure 29: Relationship of Shipping Point Node as Origin



Source: Neo4j Desktop-1.2.8, 2020

3.3.4.6. Forwarder Node and Relationships

Forwarder node has also some properties such as code, name and address data signed in as it is indicated in **Table 17**. This node helps to get information about deliveries are carried by which forwarder and creates ties for performance information in level of name of forwarder.

Table 17: The Properties of Forwarder Node

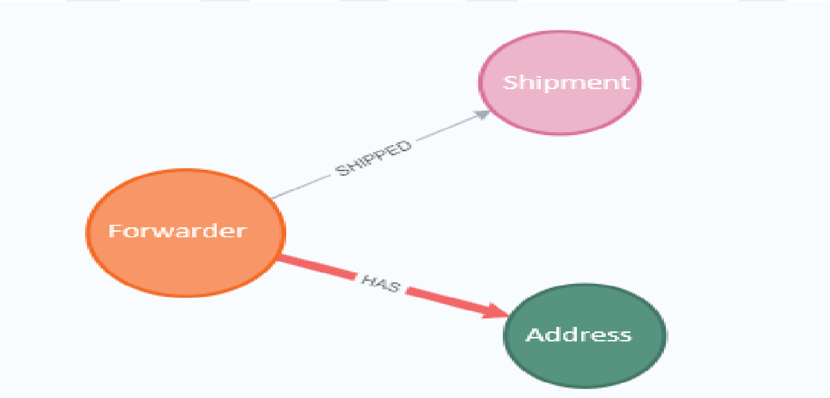
COLUMN_NAME	**ORDINAL_POSITION**
ID	1
Code	2
Address	3
Name	4

Source: Schaeffler Data Archive, 2020

There are two different relationship where forwarder node is start point. One is SHIIPPED relationship, which is name of connection with shipment node, other one is HAS relationship which is linked to address node as it is clarified in **Figure 30**. Below query provides us relationships with relevant node obviously.

match(s:Shipment)<-[h:SHIPPED]-(f:Forwarder)-[h1:HAS]->(a:Address)
RETURN f,s,a,h,h1 LIMIT 1

Figure 30: Relationships of Forwarder Node as Origin



Source: Neo4j Desktop-1.2.8, 2020

The attributes of SHIIPPED relationship are several. Since, this relationship provides us an opportunity examine logistics performance of forwarder. LPF shows logistics performance of forwarder regarding shipment number. On the other hand, LPF_LTS indicates transportation scheduled lead time and LPF_LTA gives information about shipment actual lead time as day. Additionally. LPF_DEV demonstrates deviation between actual and scheduled lead time data. Those, critical lead time information provides data about forwarder performance during logistics activities based on shipment numbers. **Table 18** shows attributes within and hidden in SHIPPED relationship between forwarder and shipment nodes in graph database in a table format.

Table 18: The Attributes of SHIPPED Relationship

COLUMN_NAME	ORDINAL_POSITION
Shipment	1
Forwarder	2
IFTMIN_Create_Date	3
LPF_Exception_Code	4
LPF_Exception_Causer_Code	5
LPF_Exception_Name	6
LPF_LTA	7
LPF_LTS	8
LPF_DEV	9
LPF	10
ID	11

Source: Schaeffler Data Archive, 2020

However, no properties included in HAS relationships between forwarder and address nodes. Since, significant information is contained in forwarder and address nodes and they are linked via HAS relationship to each other.

3.3.4.7. User Node and Relationships

User node is mentioned before with CREATED_BY relationship which is connected to delivery node. For this relationship with delivery node, user node includes following properties in **Table 19**. These properties ensure transparency for the users who are creating orders for deliveries within Schaeffler. Therefore, user node contains username (it might be different than real name of colleague), name, surname, title, department and contact information. Additionally, user node does not have any relationship where it is a start node in developed data architecture.

Table 19: The Properties of User Node

COLUMN_NAME	ORDINAL_POSITION
ID	1
UserName	2
EmployeeNumber	3
mobile	4
Email	5
Title	6
FirstName	7
LastName	8
DisplayName	9
Department	10
telephonenumber	11

Source: Schaeffler Data Archive, 2020

3.3.4.8. Incoterms Node and Relationships

One of relevant node of delivery node is incoterms node. Due to importance of incoterms where it is mentioned in chapter 1, this node is needed for logistics activities. Otherwise, risk and cost responsibilities would not be clear in case of some problems which is faced in logistics operations. For this reason, incoterms node is available and includes below properties which is demonstrated on **Table 20**. It has incoterm, incoterm description and clause type (E/F or C/D terms). Detailed information about incoterms is given in chapter 1.

Table 20: The Properties of Incoterms Node

COLUMN_NAME	ORDINAL_POSITION
ID	1
Incoterm	2
Inco_Description	3
Clause_Typ	4

Source: Schaeffler Data Archive, 2020

Incoterm node does not have any relationship where it is a start node. Related relationship with delivery node as HAS relationship is explained under the title of delivery node.

3.3.4.9. Address Node and Relationships

One of quite significant node is address node in developed data structure for this study. Code helps to define same addresses which are written different in the data. On this way, duplicated information is eliminated. Address node has place, name and label information as it is indicated on **Table 21** . Label has another level importance because it includes information about customer which is extern or intern. Schaeffler is a huge company and it also has some logistics flow for inter company shipment. These shipments defined in label intern whereas external customer determined as extern to distinguish customers.

Table 21: The Properties of Address Node

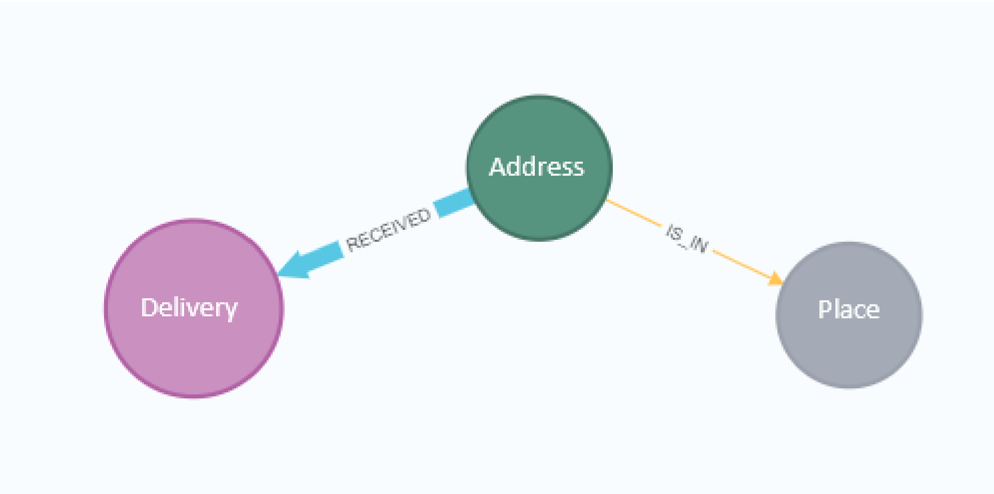
COLUMN_NAME	ORDINAL_POSITION
ID	1
Code	2
Place	3
Label	4
Name	5

Source: Schaeffler Data Archive, 2020

Figure 31 shows defined two relationships in data model where address node is start point. Address node linked to delivery node via RECEIVED relationship. It indicates address node received which deliveries. Meanwhile, IS_IN relationship ensures connection between address and place nodes. Figure 30 can be feasible with below query.

```
match(pl:Place)<-[i:IS_IN]-(a:Address)-[r:RECEIVED]->(d:Delivery)
return a,r,d,pl,i limit 1
```

Figure 31: Relationships of Address Node as Origin



Source: Neo4j Desktop-1.2.8, 2020

On **Table 22**, LD_A delivery date gives information about actual delivery date whereas LD_P delivery date shows planned delivery date. On this way delivery actual submittal date and planned date of delivery for determined address can be compared. Thus, it gives another approach for monitoring performance and catching an indicator for material flow in timely manner. Since, same material might be transported several times to same address, but it ensures a filter option. In contrast, RECEIVED relationship, IS_IN relationship does not include any attributes.

Table 22: The Attributes of RECEIVED Relationship

COLUMN_NAME	ORDINAL_POSITION
ID	1
Delivery	2
Address	3
LD_A_Delivery_Date	4
LD_P_Delivery_Date	5

Source: Schaeffler Data Archive, 2020

3.3.4.10. Place Node and Relationships

Place node insists locational information for address node. It includes place data for plant, shipping point and forwarder via address node as a bridge node. Mentioned locational data is city, postcode, latitude, longitude information as it is shown on **Table 23**. Also, country, subregion and market region properties provided connection with other nodes which will be mentioned following part of this study.

Table 23: The Properties of Place Node

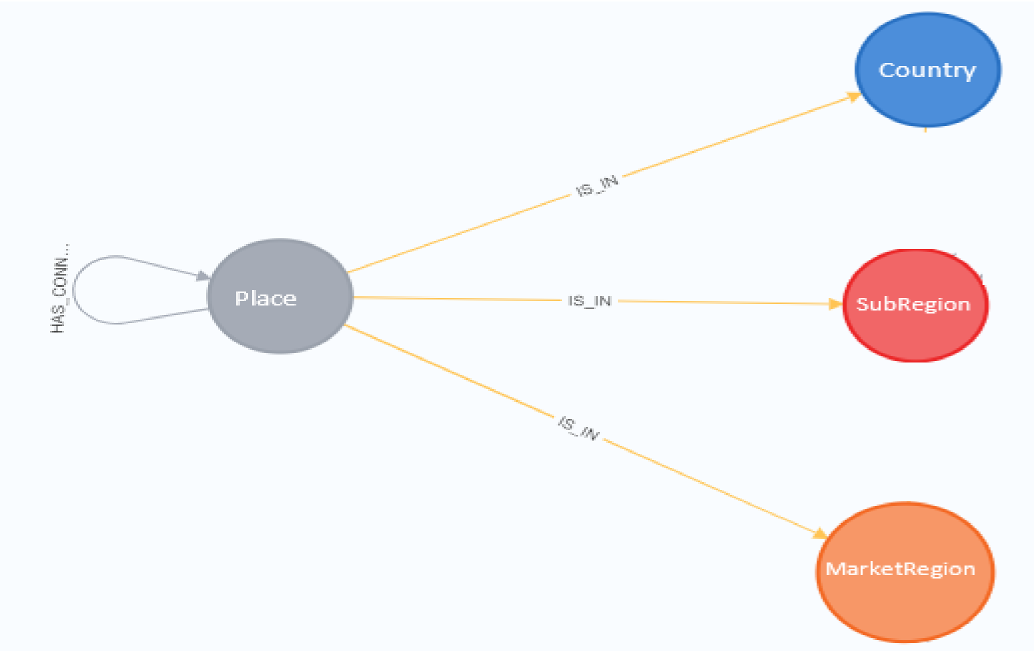
COLUMN_NAME	ORDINAL_POSITION
ID	1
City	2
Postcode	3
Latitude	4
Longitude	5
Country	6
Subregion	7
MarketRegion	8

Source: Schaeffler Data Archive, 2020

There are four relationships where place node is origin. Three of them IS_IN relationships which are linked to country, subregion and market region nodes to identify this place is in which country, subregion and market region. Another relationship is HAS_CONNECTION relationship as it is demonstrated in **Figure 32**. Since, there is physical logistics operation between nodes that's why HAS_CONNECTION relationship carries an importance. Also, below cypher query helps to get **Figure 32** in Neo4j.

```
match(pl:Place)-[r:IS_IN]->(c:Country)
match(pl)-[r1:IS_IN]->(sb:SubRegion)
match(pl)-[ r2:IS_IN]->(mr:MarketRegion)
match(pl)-[r3:HAS_CONNECTION]->(pl)
return pl,c,sb,mr,r,r1,r2,r3 limit 1
```

Figure 32: Relationships of Place Node as Origin



Source: Neo4j Desktop-1.2.8, 2020

On **Table 24**, place to place distance(km) information is included as attributes of HAS_CONNECTION relationship. It helps to find out distances between places where transportations are occurred.

Table 24: The Attributes of HAS_CONNECTION Relationship in Place Node

COLUMN_NAME	ORDINAL_POSITION
ID	1
NodeFrom	2
NodeTo	3
Distance_km	4

Source: Schaeffler Data Archive, 2020

3.3.4.11. Country Node and Relationships

Country node includes below properties, on **Table 25**. These properties are name, code of country, subregion and market region information to ensure connection with subregion and market region nodes.

Table 25: The Properties of Country Node

COLUMN_NAME	ORDINAL_POSITION
ID	1
code	2
name	3
SubRegion	4
MarketRegion	5

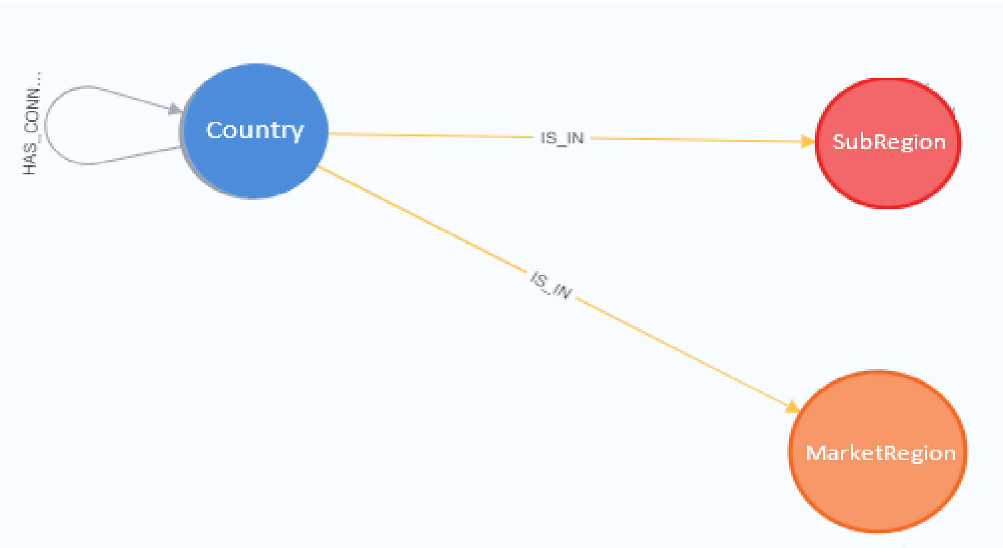
Source: Schaeffler Data Archive, 2020

Demonstrated below cypher query, brings relationships where country node is start point.

```
match(mr:MarketRegion)<-[i:IS_IN]-(c:Country)-[i1:IS_IN]-(sb:SubRegion)
match(c)-[i2:HAS_CONNECTION]-(c)
return c,sb,mr,i,i1,i2 limit 1
```

In **Figure 33**, we can see visualized version of relationships for country node. HAS_CONNECTION and IS_IN relationships are way to make connection with country node itself and subregion and market region nodes. It ensures data for which country is in which subregion and belong to which market region to make overall analysis. IS_IN relationships do not have any attributes in data model.

Figure 33: Relationships of Country Node as Origin



Source: Neo4j Desktop-1.2.8, 2020

IS_IN relationships do not have any attributes in data model. However, HAS_CONNECTION relationship within country nodes has below attributes as country to and country from to ensure transparency of country to country shipments on **Table 26**.

Table 26: Country Node and HAS_CONNECTION Relationship

COLUMN_NAME	ORDINAL_POSITION
ID	1
Country_From	2
Country_To	3

Source: Schaeffler Data Archive, 2020

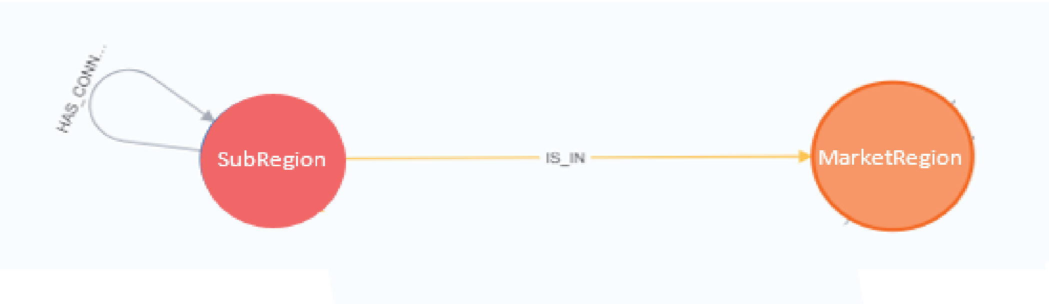
3.3.4.12. Subregion Node and Relationships

Subregion node includes title, ID and market region information as properties. Subregion node has also relationship as they are brought with below cypher query.

```
match(mr:MarketRegion)<-[i:IS_IN]-(sb:SubRegion)
match(sb)-[i1:HAS_CONNECTION]-(sb)
return sb,mr,i,i1 limit 1
```

To understand how subregion node and its relationships looks like in graph data base, we can look at **Figure 34** . Subregion has HAS_CONNECTION relationship with itself and IS_IN relationship with market region.

Figure 34: Relationships of Subregion Node as Origin



Source: Neo4j Desktop-1.2.8, 2020

On **Table 27**, shown attributes of subregion node are hidden in HAS_CONNECTION relationship of subregion node itself. It shows from subregion to another subregion information.

Table 27: Subregion Node and HAS_CONNECTION Relationship

COLUMN_NAME	ORDINAL_POSITION
id	1
SubregionFrom	2
SubRegionTo	3

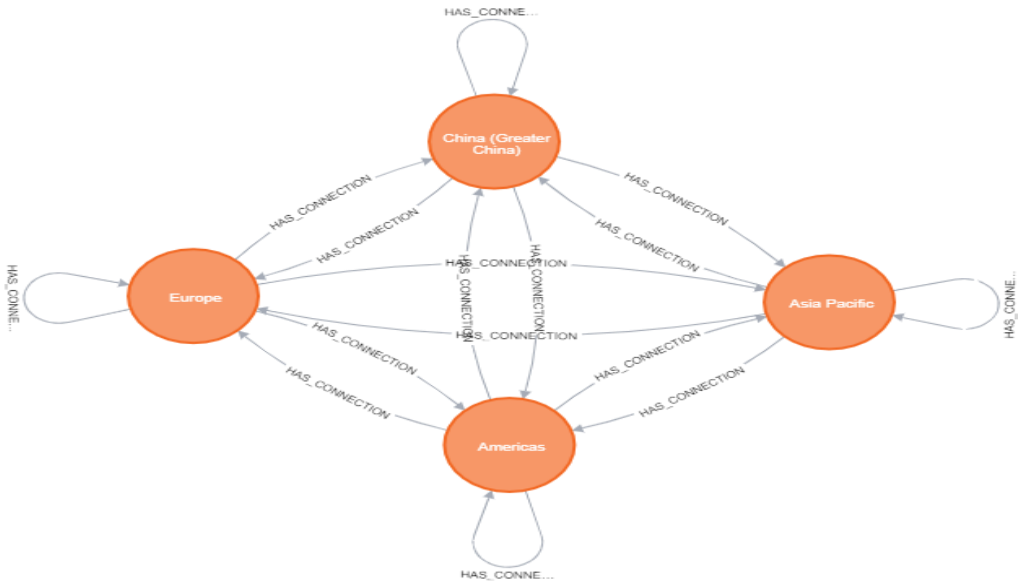
Source: Schaeffler Data Archive, 2020

3.3.4.13. Market Region Node and Relationships

Market region nodes contains only ID and title properties that's why it is not shown in table format. In

Figure 35, it is obvious to see that market regions where Schaeffler is operating in logistics level. Operated market regions are China, Europe, Asia Pacific and America and all nodes have HAS_CONNECTION relationships which are connected to each other.

Figure 35: Market Region Node with Title and Relationships



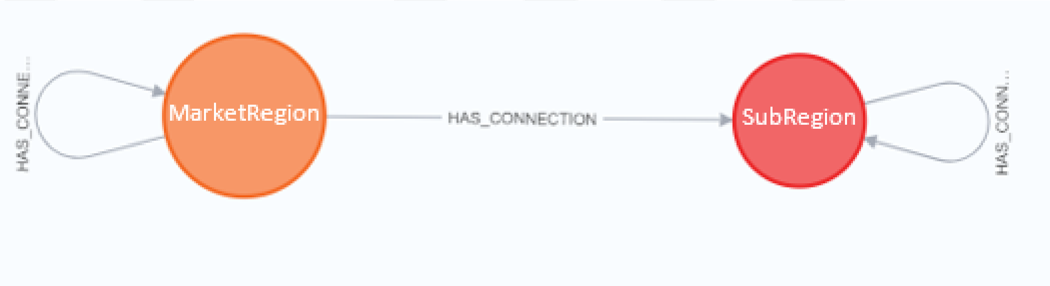
Source: Neo4j Desktop-1.2.8, 2020

To create **Figure 36** in graph database, below cypher query would be necessary in the data model in this study.

```
match(mr:MarketRegion)-[i:HAS_CONNECTION]->(sb:SubRegion)
match(mr)-[i1:HAS_CONNECTION]->(mr)
return sb,mr,i,i1 limit 1
```

Market region node is kind of end node in data structure, that's why it has HAS_CONNECTION relationship where it linked to subregion node.

Figure 36: Relationships of Market Region Node as Origin



Source: Neo4j Desktop-1.2.8, 2020

HAS_CONNECTION relationship of market region node includes data about from one market region to other market region attributes on **Table 28** as it is demonstrated.

Table 28: Market Region Node and HAS_CONNECTION Relationship

COLUMN_NAME	**ORDINAL_POSITION**
id	1
MarketRegionFrom	2
MarketRegionTo	3

Source: Schaeffler Data Archive, 2020

3.3.5. Graph Algorithms

Graph algorithms can be defined as a subset of appliances to use in graph analytics (Needham & Hodler, 2019: 3). Graph algorithms helps to analyze connected data in graph database. Since, graph algorithms ensure one of effective approach to analyze data with mathematical calculations which is functionated in specified relationships in graph data model (Needham & Hodler, 2019).

In following part of study, some graph algorithms will be examined because Neo4j is giving opportunities to use graph algorithms, such as page rank, betweenness centrality and shortest path algorithm.

3.3.5.1. Page Rank

Page Rank is one of the most significant and well-known centrality algorithms which is used in graph database. It indicates the importance of nodes and ensure a prioritization for the concerned nodes regarding its frequency on relationships. It measures influence on nodes by number of outflow connections (Needham & Hodler, 2019). As future case suggestion, it is possible to see priority level of place node in our data model, it would be helpful for prioritized location to make logistics optimization.

Mathematical formula is shown how page rank is calculated and the meaning of symbols are as below in **Figure 37**;

Figure 37: Mathematical Calculation of Page Rank for Node A

$$PR(A) = (1-d) + d\left(\frac{PR(T_1)}{C(T_1)} + \ldots + \frac{PR(T_n)}{C(T_n)}\right)$$

Source: Neo4j Graph Data Science, 2021

- d is damping factor to eliminate extreme results used in Neo4j as 0.85 (0-1)
- T_1 to T_n indicates number of node A from 1 to n
- $C(T_1)$ to $C(T_n)$ shows number of outflow relationships from node A

To understand prioritized node in logistics graph data model, page rank graph algorithm would be quite useful for future analysis. In following part another centrality algorithm will be explained which is used in Neo4j.

3.3.5.2. Betweenness Centrality

Betweenness centrality is another centrality graph algorithm and helps to understand importance of node in modelled graph network. However, there is a reason which it distinguishes betweenness centrality algorithm form page rank. Betweenness centrality is considering on being bridge node in a comparison with other nodes (Needham & Hodler, 2019). For this graph algorithm, influence is determined with inflow and outflow connections to node in graph data model regarding their shortest paths.

The betweenness centrality of a node is calculated regarding below **Figure 38*Figure 38*** where *U* is node which in between two other nodes t and *s*. All these nodes are different than each other.

Figure 38: Mathematical Calculation of Betweenness Centrality for Node U

$$B(u) = \sum_{s \neq u \neq t} \frac{p(u)}{p}$$

Source: Needham & Hodler, 2019: 93

• *U* shows node which its betweenness centrality is calculated.

• *P* indicated the total number of shortest paths between nodes *t* and *s*.

According to below query which is explained by Needham & Hodler (2019), we can get example centrality table in **Figure 39** as result and **Figure 40** is graph/visualized version of betweenness centrality algorithm result as below.

```
CALL algo.betweenness.stream(“User”, “FOLLOWS”)
YIELD nodeId, centrality
RETURN algo.getNodeById(nodeId).id AS user, centrality
ORDER BY centrality DESC
```

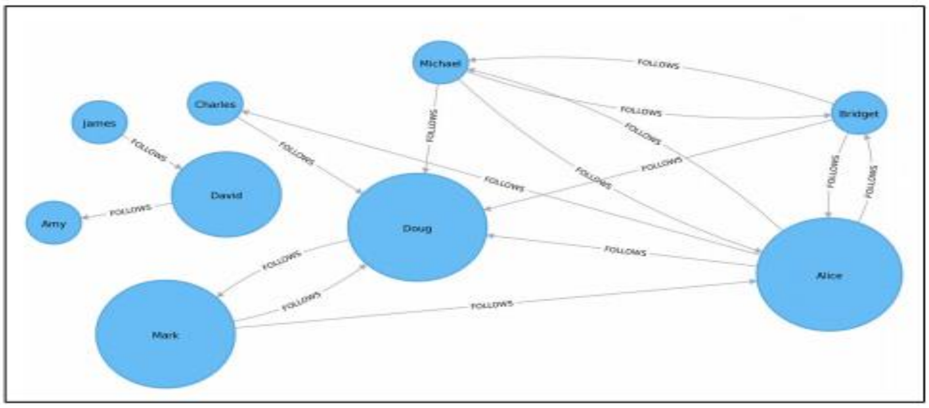
Figure 35 shows that Alice is bridge node in centrality level with 10 centrality rank regarding follows relationship with other nodes.

Figure 39: Betweenness Centrality Example in Table Format

user	centrality
Alice	10.0
Doug	7.0
Mark	7.0
David	1.0
Bridget	0.0
Charles	0.0
Michael	0.0
Amy	0.0
James	0.0

Source: Needham & Hodler, 2019: 95

Figure 40: Visualization Example of Betweenness Centrality



Source: Needham & Hodler, 2019: 96

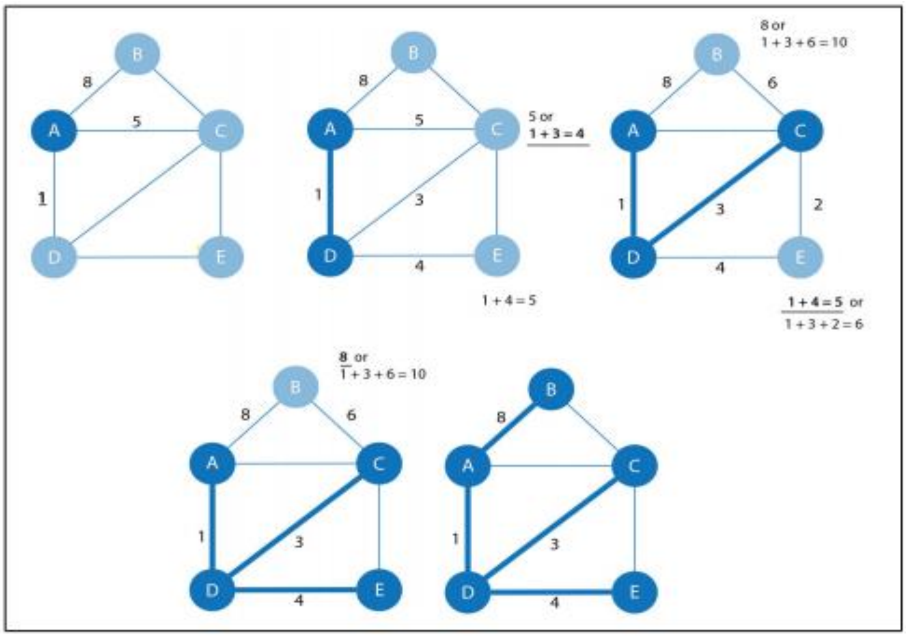
3.3.5.3. The Single Source Shortest Path Algorithm

The Single Source Shortest Path (SSSP) algorithm is the way of calculation weighted path from a start node to other nodes. SSSP became important at the same time with Dijkstra’s Shortest Path algorithm to apply on problems (Needham & Hodler, 2019: 65).

In logistics sector, shortest path algorithm can be quite helpful to understand optimal routes during physical flow of transportations.

Regarding below SSSP example A is a root node which shown in **Figure 41**. For ongoing route path smallest weighted relationships are selected. That's why connection started with firstly node D; A-D=1 and to complete all graph model C node is chosen and A-D-C= 4 paths is used due to lower weight of relationship. This process continues until there is no more node remains. Also, following **Figure 41** shows which paths are selected until enriching all nodes from node A.

Figure 41: The Steps of the Single Source Shortest Path Algorithm



Source: Needham & Hodler, 2019: 66

For detailed cypher query information about graph algorithms, Neo4j Graph Data Science website can be checked, it contains many examples which is shown in references of this study.

To ensure efficient usage of graph algorithms, we need to have wide network based on one determined relationship between start nodes and end node. However, in our data model, we focused on relationships between various nodes and their properties for outbound transports.

3.4. DATA ANALYSIS AND RESULTS

Data modelling with ERD and collected by using SAP(EP1) and SQL Server is explained in detail level under the title of nodes and relationships. In following section of study, answers for research questions will be shown, suggestion for platform selection will be examined and limitation of study will be explained.

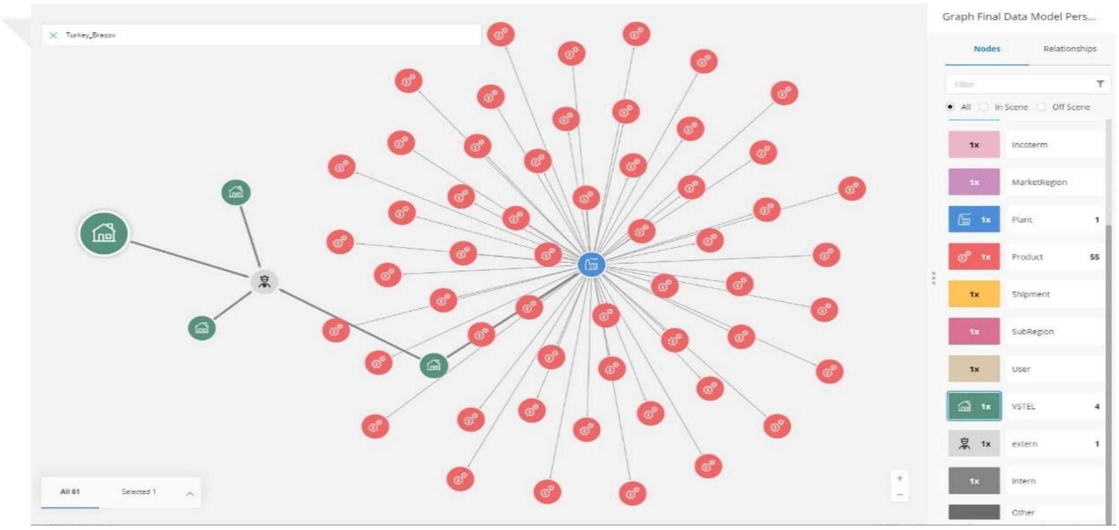
3.4.1. Results of Research Questions

First research question is about data visualization as “R_1: Is it possible to visualize outbound transports level of material flow?”. To answer this question, Neo4j Bloom is used. Neo4j Bloom is a visualization tool which is commonly used in open data source Neo4j. Due to huge data volume to eliminate chaotic results, some filters are settled in to Neo4j Bloom with a cypher query. Below cypher query helps to enrich **Figure 42** as visualized version of data. The query ensures a filter for deliveries are transported from production plant Brasov, Romania to customer which is located in Turkey. It brings deliveries for external customer in Turkey demands which final materials and these deliveries loaded in which shipping points. Also, code of these shipping points should not be in same place to catch different locations. The below query gives information on table format; however, it is not added to this study due to data privacy of the company.

```
match(sp:ShipPoint)<-[:LOADED_IN]-(d:Delivery)<-[:RECEIVED]-
(e:extern)-[:IS_IN]-(p:Place)-[:IS_IN]->(c:Country {code:"TR"})
match(d)-[:HAS]->(pr:Product)<-[:PRODUCES]-(pl:Plant)-[:HAS]
->(a:Address)-[:IS_IN]->(pl_place:Place)
-[:IS_IN]>(country_plant:Country{code:"RO"})
return distinct pr.code, sp.code,sp.name
```

Figure 42 indicates that with blue marked node is plant in Brasov, green nodes are shipping points which are used for external customers to deliver products, gray icon shows one customer in Turkey and red nodes are products which are produced in plant Brasov. It is obvious to see that products are coming to three different shipping points which are in Germany to ship out to Turkey instead of making direct shipments to Turkey. It leads to increase lead time and transport cost for the Schaeffler. However, it is hard to dedicate this kind of situation in table format data analysis and graph data base helped to detect this kind of case and visualize data. Also, one shipping point is bigger than other green shipping points, the reason of it using count regarding number of deliveries. Research question 1, showed us final material flow for one customer which is located in Turkey with starting point in plant node.

Figure 42: Visualized Material Flow from Brasov to Turkey



Source: Neo4j Desktop-1.2.8, 2020

R2: How can we have combined platform in material, delivery and shipment level? Neo4j ensured this opportunity with graph database and data volume which used is quite high. On this way shipment, delivery and material levels are combined in graph database. Previously these levels were separated in current database which is used frequently by colleagues who are working in Logistics Network Design department. SHIPPED relationship between forwarder and shipment nodes includes information about logistics forwarder performance also delivery node makes a connection with product and shipment nodes to enrich material flow in delivery level. It helps to make aggregated analysis with all levels together and eliminates one million rows limit in excel file in current database.

R_3: Is it possible to find alternative plant for the same material which is produced for the same customer? This research question is quite important to understand logistics network structure. Regarding company material flow requirement, every market region produces unique product to correspond demands in its own market. However, there might be some exceptions and it is possible to catch this kind of situations with graph database. To understand this below query written.

Firstly, labels which are included in address nodes are removed. As it is mentioned before, label of intern is used for inter-company shipment and internal receivers whereas label of extern indicates external recipients. Via below queries labels became different nodes which acts like address node in case of having same relationships. It is needed to define external receiver for the material flow.

```
Match (n:Address {label:"intern"})
call apoc.create.addLabels([id(n)],[n.label]) yield node
with node
remove node.label
return node
Match (n:Address{label:"extern"})
call apoc.create.addLabels([id(n)],[n.label]) yield node
with node
remove node.label
return node
```

In second level, via below Cypher query and starting with match clause, delivery which received by external customers are determined and filtered product which has industrial (I) division. Since, industrial customers have more unstructured network in transportation in comparison with automotive customers. Customer name, customer code, product code is brought via with clause and ordered by number of deliveries which includes products in determined deliveries. On the other hand, delivery freight payer selected as Schaeffler. Since, desire is focusing on deliveries which Schaeffler is responsible in transportation cost level. Record limited with 30 to decrease retrieving performance issues meanwhile collecting to information about receiver name which is called as WEs.

Also, match clause with place node helps to bring data about location information of extern node and at the same time extern should receive determined delivery with loaded shipping point information. Usage of delivery node makes connection to plant information via product node. In HAS relationship between delivery and product node, plant data is also available in this relationship as an attribute where it is mentioned before in **Table 13**. That's why, this plant code should be equal to plant where product is produced. Delivery node ensures connection to catch product which is shipped to determined external customer. To get locational information place node used also for plant node with connection of intern node and IS_IN and HAS relationships. Furthermore, plant type filtered as "P", it means that product producer plant. Since, **Table 4** also showed properties of plant node and here this property applied as a filter to select only real producer plant. This plant is determined as alternative plant (P1) which produces same product and shipped in another delivery but to the same customer (external receiver).

With clause helped to call locational name regarding latitude and longitude properties of place node which is indicated in **Table 23**. After naming receiver place (WElocation), plant place (PlantLocation) and other alternative plant location (Other_Plant_Location) which is producing same product, it was necessary to calculate distance kilometer data between plant and receiver location (distance_P2WE) and alternative plant and receiver distance information (distance_oP2WE). Also, these plants should be far way to each other to catch critical material flow. Therefore, difference between current plant and alternative plant should be more than 500 kilometers. To bring in table format column names are determined with return clause as code of plant, name of plant, name of other (alternative) plant, final product number, code of shipping point, name of shipping point, distance information from current plant to receiver and distance from alternative plant to receiver. Additionally, number of deliveries include in return clause with rounded total amount and ordered by sum of delivery number with limited 10 records.

```
match(p:extern)-[:RECEIVED]->(dd:Delivery)-[:HAS]-
>(pr:Product{division:'I'})
where dd.frachtzahler='Schaeffler'
with p.code as code, p.name as CustomerName, pr.code as
Product,count(distinct dd.code) as Delivery
ORDER BY Delivery DESC
LIMIT 30
with collect(code) as WEs
match(wea:Place)<-[:IS_IN]-(w:extern)-[:RECEIVED]->(d:Delivery)-
[:LOADED_IN]->(sp:ShipPoint) where w.code in WEs
match(d)-[h:HAS]->(pr)<-[:PRODUCES]-(p:Plant)-[:HAS]->(:intern)-
[:IS_IN]->(pea:Place) where h.plant=p.code
match (pea1:Place)<-[:IS_IN]-(:intern)<-[:HAS]-(p1:Plant)-[:PRODUCES]-
>(pr)<-[h1:HAS]-(d1:Delivery)<-[:RECEIVED]-(w) where h1.plant=p1.code
and d1.code <> d.code and p1.code <> p.code and p1.plantType='P'
with p,p1,w,pr,sp,d,point({ longitude: wea.longitude, latitude: wea.latitude })
AS WElocation, point({ longitude: pea.longitude, latitude: pea.latitude }) AS
PlantLocation,point({ longitude: pea1.longitude, latitude: pea1.latitude }) as
Other_Plant_Location
with p.code as plant,p.name as plantname,p1.name as other_plant,d.code as
lieferung, d.we as Customer,d.weName as Customername,pr.code as
product,sp.code as SP,sp.name as SPname,round(distance(PlantLocation,
WElocation))/1000.0as distance_P2WE,round(distance(Other_Plan_Location,
WElocation))/ 1000.0 as distance_oP2WE
where distance_P2WE-distance_oP2WE>500
return
plant,plantname,other_plant,Customer,Customername,product,SP,SPname,dis
tance_P2WE,distance_oP2WE,count(lieferung) as count_delivery
order by count_delivery desc
limit 10
```

Due to data privacy of company, it is not allowed to share detailed information especially customer and material level data in detail. That's why, result will be shared in table format with overall picture for final materials A, B, and C logistics flow. Each material is produced for same customer based on defined requirements Thus, criteria of different type of demand requirement is eliminated. The weight assumption for one-year dispatches is 20 tons for each material. **Table 29** shows status of material flows and alternative plant locations regarding decreased distance information. In some deliveries, material A, B, C dispatched from production plants where are in India and Vietnam for the customers which are in market region Europe. Lead time for sea transport mode is approximately 60 days from ports in India and Vietnam to port of Hamburg. However, same materials are also producing in production plants where in Germany and Slovakia. There is a rule within company as every market region produces unique final material. On this way demands should be delivered from plants which are in same market region with customers. The lead time can be decreased 90% for determined materials via selection of road transport mode instead of sea. Thus, increased time when capital on the way can be avoidable. Also risks might decreased during the logistics operations. Also, transportation cost can be decreased 30% for each material. On the other hand, deliveries are going to shipping points unnecessarily. It gives an opportunity to use warehouse for stocking meanwhile increases lead time, transportation cost and work force.

Table 29: Current Versus Alternative Transport Path

	Country of Plant		**Total Distance (km) to Receiver**			
Material	**Current**	**Alternative**	**Current**	**Alternative**	**Shipping Point Country**	**County of Customer**
A	India	Germany	13.220	290	Germany	Germany
B	India	Slovakia	13.130	750	Germany	Germany
C	Vietnam	Slovakia	21.470	2.720	Germany	Spain

Source: Neo4j Desktop-1.2.8, 2020

R_4: Is it feasible to find shortest path between plants and end customers?

Unfortunately, it is not possible to apply any graph algorithms with current data structure. Traceability is not possible after delivery arrives in one loading point and delivery number is changing at every new dispatching place. That's why, data is not connected to catch shortest path in graph database. In future, project regarding updating data model and including new SAP tables, shortest path can be calculated in logistics network.

3.4.2. Limitations of Study

In this study, many obstacles are occurred during the data modelling. One and the most important limitation raised up from data quality. During the data collection, it is detected that geolocation of plants was not available or some plants and shipping points were already shut down. A new geo coordination list was prepared, and the problem was solved in this study, but data was not clean in the beginning of project. When the material came out from for the production line, material number has 9 digits, meanwhile it becomes 13 digits after packaging process and 15 digits material number indicates the final material number which is ready to sell final customer. This knowledge was gained after the project has been already started, it leaded to update data model to include only final material and time consumption.

Data diversity and complexity is another limitation encountered with. The amount of data was quite huge and created bunch of data complexity. For instance, 135 Schaeffler entities names defined as Schaeffler Technologies AG & Co. KG in the data. However, they are different than each other and it was a necessity differentiate them with additional variables. Also, Schaeffler has intercompany shipments. That's why, a plant or shipping point might act like a customer. To prevent making a wrong data analysis, customer node with a label called as "intern" and separated from real production plant node. On the other hand, duplicates in address node is detected. Address ID number were used because written format of address information was also quite complicated, but even id numbers had some duplicates. Also, for a time the delivery number could have different shipment numbers also one shipment would have various delivery numbers. This situation also increased data complexity and made analyzing the data harder. Another diversity reason raised from using different data sources in Schaeffler. For example, invoice verification department is using separated data bank to store data and transportation cost information is getting from this source basically, but data availability was not high and it decreased reliability in the study.

Performance issues can be part of the limitation of this study. Transferring all one-year outbound shipments from SQL server is hard. Since we need to have a single table for each node which is created on Neo4j database. All titles must match in logistics network design database and SQL in the way of writing. Also, due to large amount of data leaded to data crush due to having not enough memory. Central Processing Unit (CPU) was limited with free trial version in Neo4j and ETL tool created additional indexes itself and it caused performance issue in timely manner.

Information Technologies (IT) department sets some security restrictions which are strict, that’s why during the testing of a new application got also faced with some obstacles.

CONCLUSION

In this study, a proof of concept defined based on graph database and shown visualization and optimization opportunities are shown within the big data. Graph database is quite famous in recent years to analyze huge amount of data. This study proved that how graph database is capable to analyze and manipulate the large data sets. On this way, it brought another and modern approach for Schaeffler to store data in contrast with the traditional way of data management.

Data migration successfully has been made form relational database to graph database for better understanding on logistics activities. The new data model gave opportunity to enrich combined level data structure meanwhile data reliability has been increased. Thus material, shipment and delivery level data met in one combined platform. Also, reporting limitations due to number of records is eliminated in logistics network design database which is used currently by logistics network design department.

This study contributed transparency on material flow for the outbound shipments in all over the world in a practical approach. Regarding production plant where is in Brasov is sending to materials indirectly; firstly, to shipping points which are in Germany and then to final external customer where is in Turkey. Visualized data structure ensured optimization possibilities based on lead time and transportation costs for these outbound shipments. On the other hand, for optimal transportation flow, alternative plants which are locationally closer to end receivers are defined to answer customers' demands and for three final materials 90% lead time and 30% logistics' cost reduction is ensured. As a result, three research questions are proved and the single shortest path algorithm could not be applied due to current data structure. Since, traceability is not possible due to changes of delivery numbers in every new dispatches.

In general overview, route optimizations can be easily defined regarding unnecessarily transportation path of materials. Thus, lead time and transportation cost can be decreased, optimized and qualified logistics network can be designed. Meanwhile this study contributes another aspect to the current literature about usage of graph database in logistics aspect with highly volumed data.

The project will be continuing in logistics network design department in Schaeffler, in data aggregation level. Next stage will be defined under the concern of budgeting and other possibilities about graph database. Benefits of graph database is clearly defined but which tool will be used is under the investigation regarding bench marking. The selection of tool might be unclear; however, graph database will be demanded to use in future projects for the big data analyses.

REFERENCES

Abushaikha, I., Salhieh, L., & Neil, T. (2018). Improving Distribution and Business Performance Through Lean Warehousing. *International Journal of Retail & Distribution Management, Vol.46, No.8*, pp. 780-800.

AlShaer, M., Taher, Y., Haque, R., Hacid, M.-S., & Dbouk, M. (2019). IBRIDIA: A Hybrid Solution for Processing Big Logistics Data. *Future Generation Computer Systems, Vol.97*, pp. 792-804.

Bag, S., Yadav, G., Wood, L. C., Dhamija, P., & Joshi, S. (2020). Industry 4.0 and the Circular Economy: Resource Melioration in Logistics. *Resources Policy, Vol.68*, pp. 1-16.

Balzer, C., Oktavian, R., Zandi, M., Jimenez, D. F., & Moghadam, P. Z. (2020, November 13). Wiz: A Web-Based Tool for Interactive Visualization of Big Data. *Patterns, Vol.1, Issue.8*, pp. 1-7.

BBC. (2021). Egypt's Suez Canal Blocked by Huge Container Ship. *BBC News*.

Bell, P. C., & Chen, J. (2015). Close Integration of Pricing and Supply Chain Decisions Has Strategic As Well As Operations Level Benefits. *Annals of Operations Research*, pp. 77-93.

Bellman, R., & Cooke, K. (1969). The Konigsberg Bridges Problem Generalized. *Journal of Mathematical Analysis and Applications, Vol.25, Issue.1*, 1-7.

Bouaziz, S., Nabli, A., & Gargouri, F. (2019). Design a Data Warehouse Schema from Document-Oriented Database. *Procedia Computer Science, Vol.159*, 221-230.

Bowersox, D. J., & Calantone, R. J. (1998). Executive Insights: Global Logistics. *Journal of International Marketing Vol. 6, No. 4*, pp. 83-93.

Chiu, Y., Lin, P.-C., & Hsu, H.-H. (2011, December 8). Considering Third-Party Logistics Providers in Reverse Logistics. *Journal of the Chinese Institute of Industrial Engineers, Vol.28, No.7*, pp. 512-520.

Chopra, A. (2020, June 13). Conceptual Framework of IoT for Transport Logistics An Approach to Connecting Material Flow and IT in Self-Directing Collaborating Logistics Progressions. *International Journal of System Assurance Engineering and Management*, pp. 258-266.

Chow, G., Heaver, T. D., & Henriksson, L. E. (1994). Logistics Performance: Definition and Measurement. *International Journal of Physical Distribution & Logistics Management, Vol.24, No.1*, pp. 17-28.

Committees, I. N. (2020). Incoterms 2020. (pp. 1-18). London: International Chamber of Commerce.

Cremona, L., Ravarini, A., & Brocke, J. V. (2018). Visualizing Big Data: The Impact on Sense-Making and Decision-Making. In F. Cabitza, B. Carlo, & M. Magni, *Organizing for the Digital World: IT for Individuals, Communities and Societies* (pp. 23-31). Milan: Springer.

Czerepicki, A. (2016). Application of Graph Databases for Transport Purposes. *Bulletin of the Polish Academy of Sciences Technical Sciences, Vol.64, No.3*, 457-466.

DB-Engines Ranking. (2021). Retrieved from https://db-engines.com/en/ranking_trend/graph+dbms

Depeau, J. (2019, May 02). *Neo4j Blog*. Retrieved from Going Underground: Graphing and Pathfinding London Tube Lines: https://neo4j.com/blog/going-underground-graphing-pathfinding-london-tube-lines/

Dyson, R. G. (2004). Strategic Development and SWOT Analysis at the University of Warwick. *European Journal of Operational Research* , pp. 631-640.

Fan, T., Pan, Q., Pan, F., Zhou, W., & Chen, J. (2019, June). Intelligent logistics integration of internal and external transportation with separation mode. *Transportation Research Part E,* p. 1.

Farahani, R. Z., Rezapour, S., & Kardar, L. (2011). *Logistics Operations and Management Concepts and Models.* Elsevier Insights-Inc.

Funke, T., & Becker, T. (2020). Complex Networks of Material Flow in Manufacturing and Logistics: Modeling, Analysis, and Prediction Using Stochastic Block Models. *Journal of Manufacturing Systems, Vol.56*, pp. 296-311.

Gripsrud, G., Jahre, M., & Persson, G. (2006, September 1). Supply Chain Management-Back To The Future? *International Journal of Physical Distribution & Logistics Management, Vol.36, Issue.8,* pp. 643-659.

Gundlach, G. T., Bolumole, Y. A., Eltantawy, R. A., & Frankel, R. (2006, December 1). The Changing Landscape of Supply Chain Management, Marketing Channels of Distribution, Logistics and Purchasing. *Journal of Business & Industrial Marketing, Vol.21, Issue7,* pp. 428-438.

Holcomb, M. C., Liao-Troth, S., & Manrodt, K. B. (2014). A Shift in Fundamentals: The Changing Direction in Logistics and Transportation Management. *Transportation Journal, Vol. 53, No. 4,* pp. 516-533.

Holmstrom, J. B., & Aavikko, P. (1994). Achieving A Management Breakthrough in Inbound Logistics by Improving The Efficacy of Operational Decisions. *Production and Inventory Management Journal, Vol. 35, Issue. 3*, pp. 1-8.

Holzschuher, F., & Peinl, R. (2015). Queying a Graph Database- Language Selection and Performance Considerations. *Journal of Computer and System Sciences, Vol.82,* 45-68.

Hou, H., Chaudhry, S., Chen, Y., & Hu, M. (2017, July). Physical Distribution, Logistics, Supply Chain Management, and the Material Flow Theory: A Historical Perspective. *Information Technology and Management, Vol.18, Issue.2*, pp. 107-117.

Hsiao, H., van der Vorst, J., Kemp, R., & Omta, S. (. (2010, June). Developing A Decision-Making Framework For Levels of Logistics Outsourcing in Food Supply Chain Networks. *International Journal of Physical Distribution & Logistics Management Vol.40, No.5*, pp. 395-414.

Hurtado, P. A., Dorneles, C., & Frazzon, E. (2019). Big Data Application for E-Commerce's Logistics: A Research Assessment and Conceptual Model. *International Federation of Automatic Control* (pp. 838-843). Elseiver Ltd.

IMF. (2020). *World Economic Outlook Update*. International Monetary Fund.

Ivanov, D., & Das, A. (2020). Coronavirus (COVID-19/SARS-CoV-2) and Supply Chain Resilience: A Research Note. *International Journal of Integrated Supply Management, Vol.13, No.1*, pp. 90-102.

Kembro, J. H., Andreas, N., & Eriksson, E. (2018, October 1). Adapting Warehouse Operations and Design to Omni-Channel Logistics: A literature Review and Research Agenda. *International Journal of Physical Distribution & Logistics Management, Vol. 48, Issue.9*, pp. 890-912.

Kemper, C. (2015). *Begining Neo4j*. New York: Springer Science Business Media.

Kengpol, A., Tuammee, S., & Tuominen, M. (2014, November 4). The Development of a Framework for Route Selection in Multimodal Transportation. *The International Journal of Logistics Management, Vol.25, No.3*, pp. 581-610.

Kim, G.-S., Lee, S.-W., Seo, Y.-J., & Kim, A.-R. (2020, January 23). Multimodal Transportation via TSR for Effective Northern Logistics: Perspectives of Korean Logistics Companies. *Maritime Business Review, Vol.5, No.3*, pp. 291-308.

Klein, H. S., & Morschett, D. (2006, May). The Relationship between Marketing Performance, Logistics Performance and Company Performance for Retail Companies. *International Review of Retail, Distribution and Consumer Research, Vol.16, No.2*, pp. 277-296.

Kücükoglu, I., & Öztürk, N. (2017). Two-stage Optimisation Method for Material Flow and Allocation Management in Cross-docking Networks. *International Journal of Production Research, Vol.55, No.2*, pp. 410-429.

Lambert, D. M., & Stock, J. R. (1982). Strategic Planning For Physical Distribution. *Journal of Business Logistics, Vol.3, No.2*, pp. 26-46.

Lecavalier, J. (2016). Logistics The First with The Most. In *The Rule of Logistics* (p. 32). University of Minnesota Press.

Lee, M.-K., & Elsayed, E. A. (2005, May 1). Optimization of Warehouse Storage Capacity Under A Dedicated Storage Policy. *International Journal of Production Reserach, Vol.43, No.9*, pp. 1785-1805.

Li, D., Shishank, S., Ruyter, A. D., Ali, S., Bailey, D., Hearne, D., & Salh, S. (2019). Brexit and the Auto Industry: Understanding the Issues for Supply Chain

Management. In A. D. Ruyter, & B. Nielsen, *Brexit Negotiation After Article 50: Assesing Process, Progress and Impact* (pp. 173-191).

Loske, D. (2020). The Impact of COVID-19 on Transport Volume and Freight Capacity Dynamics: An Empirical Analysis in German Food Retail logistics. pp. 1-6.

Lummus, R. R., Krumwiede, D. W., & Vokurka, R. J. (2001). The Relationship Of Logistics To Supply Chain Management:Developing A Common Industry Definition. *Industrial Management & Data Systems*, pp. 426-431.

Mariani, G. (2017). A SWOT Analysis of Acquisition Strategy. In G. Mariani, *M&A and Value Creation: A SWOT Analysis* (pp. 39-78). Italy: G. Giappichelli.

Marti, L., Puertas, R., & Garcia, L. (2014). The Importance of the Logistics Performance Index in International Trade. *Applied Economics, Vol.46, No.24*, pp. 2982-2992.

Nakamura, K., Yamada, T., & Tan, K. H. (2019). The Impact of Brexit on Designing a Material-based Global Supply Chain Network for Asian Manufacturers. *Management of Environmental Quality, Vol.30, Issue.5*, pp. 980-1000.

Needham, M., & Hodler, A. E. (2019). *Graph Algorithms: Practical Examples in Apache Spark & Neo4j*. USA: O'Reilly Media, Inc.

Neo4j Customers. (2020). Retrieved from https://neo4j.com/customers/

Neo4j Desktop-1.2.8. (2020). Nürnberg, Germany.

Neo4j Desktop-1.2.8-ETL Tool. (2020). Nürnberg, Germany.

Neo4j Graph Data Science. (2021). Retrieved from https://neo4j.com/docs/graph-data-science/current/algorithms/page-rank/

Nobari, A., & Kheirkah, A. (2017). Integrated and Dynamic Design of Sustainable Closed-Loop Supply Chain Network Considering Pricing. *Transactions E: Industrial Engineering, Vol.25, No.1*, pp. 410-430.

Onstein, A. T., Tavasszy, L. A., & Van Damme, D. A. (2019). Factors Determining Distribution Structure Decisions in Logistics: A Literature Review and Research Agenda. *Transport Reviews, Vol.39, No.2*, pp. 243-260.

Panzarino, O. (2014). *Learning Cypher*. Birmingham, UK: Packt Publishing.

Patowary, K. (2018, August 7). *Amusing Planet*. Retrieved from What Mathematics Has to Do With The Seven Bridges of Königsberg: https://www.amusingplanet.com/2018/08/the-seven-bridges-of-konigsberg.html

Pilar, L., Pokorna, J., Polakova, J., & Selby, R. (2015). SWOT Analysis: Tool to Identify Students' Perception of Study Programs. *12th International Conference on Efficiency and Responsibility in Education 2015,Vol.12* (pp. 452-459). Prague: Researchgate.

Porche, I. R., Wilson, B., Johnson, E.-E., Tierney, S., & Saltzman, E. (2014). Big Data: Challenges and Opportunities. In I. R. Porche, B. Wilson, E.-E. Johnson, S. Tierney, & E. Saltzman, *Data Flood: Helping the Navy Address the Rising Tide of Sensor Information* (pp. 1-6). RAND Corporation.

Robinson, I., Webber, J., & Emil, E. (2015). *Graph Databases: New Opportunities For Connected Data.* United States of America: O`Reilly Media Inc.

Rutner, S. M., & Langley, C. J. (2000, July 1). Logistics Value: Definition, Process and Measurement. *The International Journal of Logistics Management*, p. 73.

Sandberg, E. (2011, October 1). Business Trends and Logistics Challenges in the Western European Retail Industry. *Journal of Business and Retail Management Research, Vol.6, Issue,1*, pp. 1-14.

Sanders, N., & Swink, M. (2020, March). Digital Supply Chain Transformation: Visualizing The Possibilities. *Logistics Management, Vol.59, Issue.3,* pp. 42-53.

SAP Tables. (2021). Retrieved from LIKP SAP SD Document: Delivery Header Data Table and Data: https://www.se80.co.uk/saptables/l/likp/likp.htm

Schaeffler. (2019). *Annual Report.* Herzogenaurach.

Schaeffler. (2019). *SCM & Logistics Europe.* Herzogenaurach.

Schaeffler. (2020). *Annual Report.* Herzogenaurach.

Schaeffler. (2020). *Financial Statements Annual Report.* Herzogenaurach.

(2020). *Schaeffler Data Archive.* Herzogenaurach: Schaeffler.

Selviaridis, K., & Norrman, A. (2015, July). Performance-based Contracting for Advanced Logistics Services: Challanges in its adoption, design and management. *International Journal of Physical Distribution & Logistics Management, Vol.45, Issue.6*, pp. 592-617.

Sink, H. L., Langley, C. J., & Gibson, B. J. (1996, April 1). Buyer Observations of The US Third-Party Logistics Market. *International Journal of Physical Distribution & Logistics Management, Vol.26, No.3*, pp. 38-46.

Skjoett-Larsen, T. (1999, June). Third Party Logistics- From An Interorganizational Point of View. *International Journal of Physical Distribution & Logistics, Vol.30, No.2*, pp. 112-127.

Slater, A. G. (1979, January 1). *International Journal of Physical Distribution & Materials Management*, *Vol.9, No.4*, pp. 183-202.

Sobczak, P., Stawiarska, E., Olah, J., Popp, J., & Kliestik, T. (2018, October 16). Logistics Management of the Rail Connections Using Graph Theory: the Case of a Public Transportation Company on the Example of Koleje Dolnosaskie S.A. *Engineering Management in Production and Services*, pp. 7-22.

Solomon, C. (n.d.). *SAP Table Relations version 1.0.0*. Retrieved from https://www.academia.edu/24141795/SAP_Table_Relations_version_1.0.0_Christopher_Solomon_with_contributions_by_various_SAP_Professionals

SteadieSeifi, M. (2011). Logistics Strategic Decisions. In R. Z. Farahani, S. Rezapour, & L. Kardar, *Logistics Operations and Management: Concepts and Models* (pp. 43-53).

Stefansson, G. (2006, February 1). Collaborative Logistics Management and The Role of Third-Party Service Providers. *International Journal of Physical Distribution & Logistics*, pp. 76-92.

Vasiliauskas, A. V. (2010, October 27). Modelling Performance of Raiways Nodes as Intermodal Terminals,Vol.21,No.3. *Transport*, pp. 155-159.

WALCOM. (2019). Retrieved from International Conference and Workshops on Algorithms and Computation: http://www.walcom-conference.org/walcom2019.php

Wallden, M., & Ozkan, A. (2016). A Graph Database Management System for a Logistics-Related Service. *Examensarbete Inom Teknik, Grundniva*, pp. 1-94.

Wang, J. W., & Wu, Y. (2019). A Continuous Approximation Approach Based on Regular Hexagon Partition For The Facility Location Problem Under Disruption Risk. *Hindawi Complexity*, pp. 1-12.

Witte, H. (1980). On The Quality of Transportation Facilities. *International Journal of Transport Economics Vol.7, No.2*, pp. 239-247.

APPENDIX

Appendix 1: Codding

Material Flow from Plant in Brasov to Customer in Turkey

```
match(sp:ShipPoint)<-[:LOADED_IN]-(d:Delivery)<-[:RECEIVED]-(e:extern)-
[:IS_IN]-(p:Place)-[:IS_IN]->(c:Country {code:"TR"})
match(d)-[:HAS]->(pr:Product)<-[:PRODUCES]-(pl:Plant)-[:HAS]
->(a:Address)-[:IS_IN]->(pl_place:Place)
-[:IS_IN]>(country_plant:Country{code:"RO"})
return distinct pr.code, sp.code,sp.name
```

Finding Alternative Plant Instead of Current Production Plant for Same Material & Customer

```
Match (n:Address {label:"intern"})
call apoc.create.addLabels([id(n)],[n.label]) yield node
with node
remove node.label
return node
Match (n:Address{label:"extern"})
call apoc.create.addLabels([id(n)],[n.label]) yield node
with node
remove node.label
return node
match(p:extern)-[:RECEIVED]->(dd:Delivery)-[:HAS]->(pr:Product{division:'I'})
where dd.frachtzahler='Schaeffler'
with p.code as code, p.name as CustomerName, pr.code as Product,count(distinct
dd.code) as Delivery
ORDER BY Delivery DESC
LIMIT 30
with collect(code) as WEs
```

```
match(wea:Place)<-[:IS_IN]-(w:extern)-[:RECEIVED]->(d:Delivery)-
[:LOADED_IN]->(sp:ShipPoint) where w.code in WEs
match (d)-[h:HAS]->(pr)<-[:PRODUCES]-(p:Plant)-[:HAS]->(:intern)-[:IS_IN]-
>(pea:Place) where h.plant=p.code
match (pea1:Place)<-[:IS_IN]-(:intern)<-[:HAS]-(p1:Plant)-[:PRODUCES]->(pr)<-
[h1:HAS]-(d1:Delivery)<-[:RECEIVED]-(w) where h1.plant=p1.code and d1.code
<> d.code and p1.code <> p.code and p1.plantType='P'
with p,p1,w,pr,sp,d,point({ longitude: wea.longitude, latitude: wea.latitude }) AS
WElocation, point({ longitude: pea.longitude, latitude: pea.latitude }) AS
PlantLocation,point({ longitude: pea1.longitude, latitude: pea1.latitude }) as
Other_Plant_Location
with p.code as plant,p.name as plantname,p1.name as other_plant,d.code as lieferung,
d.we as Customer,d.weName as Customername,pr.code as product,sp.code as
SP,sp.name as SPname,round(distance(PlantLocation, WElocation))/ 1000.0 as
distance_P2WE,round(distance(Other_Plan_Location, WElocation))/ 1000.0 as
distance_oP2WE
where distance_P2WE-distance_oP2WE>500
return
plant,plantname,other_plant,Customer,Customername,product,SP,SPname,distance_
P2WE,distance_oP2WE,count(lieferung) as count_delivery
order by count_delivery desc
limit 10
```