**MINISTRY OF EDUCATION OF THE REPUBLIC OF AZERBAIJAN**

**AZERBAIJAN STATE OIL AND INDUSTRY UNIVERSITY**

**“COMPUTER ENGINEERING” DEPARTMENT**

**Faculty: “Information Technologies and Management”**

**Department: “Computer engineering”**

**Specialty: “050655” Information Technology**

**Group: 682.18E**

**GRADUATION WORK**

**Subject**: Research on the application of big data technologies in the transport sector

**Student:**  Ismayilzada I.R.

**Supervisor:** assoc. prof. Abdullayev V.H.

**Compliance assessment:** assoc. prof. Asgarova B.H.

**Head of the department:** assoc. prof. Ragimova N.A.

**BAKU 2022**

**The graduation work was checked by the Anti-plagiarism System:**

**D.Sc.Tech. dosent Damirova C.R**

**ABSTRACT**

**Subject: “Research on the application of big data technologies in the transport sector”**

As evidenced by numerous global projects, big data is becoming a research focus in the transportation and logistics industry. Data will be generated in significant quantities by new technologies. They generate massive amounts of data that will significantly impact the design and implementation of transportation systems, making them safer, more efficient, and profitable. As a result, research into big data analytics in transportation is a burgeoning field. This paper begins by reviewing the history and characteristics of big data and transportation systems. Next, the big data analytics framework is discussed, including an overview of data sources and sampling techniques, data analytics techniques and systems, and big data analytics software categories. Because users share personal geo-localised data and content via their devices and computers, privacy is likely the most stringent security requirement imposed by Big Data-enabled transportation applications. As a result, we will discuss the unsolved threats to the transportation industry’s private data collection. Finally, this paper discusses some of the unsolved problems associated with big data technologies.

**Keywords:** big data, database, IoT, AI, transportation, logistics

**XÜLASƏ**

**Mövzu: “Nəqliyyat sektorunda Big Data texnologiyasının tətbiqinin araşdırılması”**

Çoxsaylı qlobal layihələrin sübut etdiyi kimi, nəqliyyat və logistika sənayesində böyük verilənlər tədqiqat mərkəzinə çevrilir. Böyük miqdarda məlumat yeni texnologiyalar tərəfindən yaradılacaq. Onlar nəqliyyat sistemlərinin dizaynına və tətbiqinə əhəmiyyətli dərəcədə təsir edəcək, onları daha təhlükəsiz, daha səmərəli və gəlirli edəcək böyük həcmdə məlumat yaradırlar. Nəticə etibarilə, nəqliyyatda böyük verilənlərin analitikasına dair tədqiqatlar inkişaf edən bir sahədir. Bu məqalədə ilk olaraq böyük verilənlərin və nəqliyyat sistemlərinin tarixi və xüsusiyyətləri nəzərdən keçiriləcək. Bundan sonra, məlumat mənbələri və toplama metodlarının xülasəsi, məlumat analitikası üsulları və platformaları və böyük verilənlər analitikası tətbiqi kateqoriyaları daxil olmaqla, böyük verilənlərin analitikası üçün çərçivə müzakirə olunacaq. İistifadəçilər öz cihazları və kompüterləri vasitəsilə şəxsi geo-lokallaşdırılmış məlumatları və məzmunu paylaşdıqlarına görə, məxfilik, çox güman ki, Big Data ilə işləyən nəqliyyat proqramları tərəfindən qoyulan ən ciddi təhlükəsizlik tələbidir. Nəticədə, biz nəqliyyat sənayesinin üzləşdiyi özəl məlumatların toplanması üçün həll edilməmiş çətinlikləri müzakirə edəcəyik. Nəhayət, bu məqalə böyük verilənlər texnologiyaları ilə bağlı bəzi həll edilməmiş problemləri müzakirə ediləcək.

**Açar sözlər:** böyük verilənlər, verilənlər bazası, IoT, AI, nəqliyyat, logistika

**РЕФЕРАТ**

**Тема: «** **Исследования по применению технологий больших данных в транспортной сфере»**

В сфере транспорта и логистики большие данные становятся предметом исследований, о чем свидетельствуют многочисленные глобальные проекты. Новые технологии будут генерировать большое количество данных. Они генерируют огромные объемы данных, которые существенно повлияют на проектирование и внедрение транспортных систем, делая их более безопасными, эффективными и прибыльными. В результате исследования в области аналитики больших данных на транспорте становятся бурно развивающейся областью. История и характеристики больших данных и транспортных систем впервые рассматриваются в этой статье. После этого обсуждается структура для анализа больших данных, включая сводку источников данных и методов сбора, методов и платформ анализа данных, а также категорий приложений для анализа больших данных. Поскольку пользователи обмениваются личными геолокализованными данными и контентом через свои устройства и компьютеры, конфиденциальность, вероятно, является самым строгим требованием безопасности, предъявляемым транспортными приложениями с поддержкой больших данных. В результате мы обсудим нерешенные угрозы сбору частных данных, с которыми сталкивается транспортная отрасль. Наконец, в этой статье обсуждаются некоторые из нерешенных проблем, связанных с технологиями больших данных.

**Ключевые слова:** большие данные, база данных, Интернет вещей, ИИ, транспорт, логистика.

CONTENTS

[CHAPTER 1. INTRODUCTION TO BIG DATA AND TRANSPORT 9](#_Toc105124470)

[1.1. Literature Review 9](#_Toc105124471)

[1.2. Transportation Industry Challenges & Opportunities 11](#_Toc105124472)

[1.3. The Impact of the IoT, Big Data and AI 13](#_Toc105124473)

[1.4. The Importance of Big Data Analytics in Transportation and Logistics 15](#_Toc105124474)

[CHAPTER 2. BIG DATA FEATURES, SCALES AND DATA SOURCES 20](#_Toc105124475)

[2.1. Big Data’s 3V 20](#_Toc105124476)

[2.2. Data Collection 23](#_Toc105124477)

[2.3. Real-time data analytics 27](#_Toc105124478)

[2.4. Security and privacy 29](#_Toc105124479)

[CHAPTER 3. ITS AND RELATED APPLICATIONS 31](#_Toc105124480)

[3.1. Intelligent transportation systems in big data 31](#_Toc105124481)

[3.2. Applications and related projects 32](#_Toc105124482)

[3.3. Open challenges 33](#_Toc105124483)

[CONCLUSION 36](#_Toc105124484)

[REFERENCES 38](#_Toc105124485)

[APPENDIX 41](#_Toc105124486)

**INTRODUCTION**

**The actuality of the work:**

This research shows how big data in transportation can assist with complex functions like tracking operations, evaluating performance, predicting outcomes, and other critical transportation and logistics operations. In this thesis, we will be looking through the changes to the transportation sector regarding Big Data. Most of these changes concern the collection, generation, storage and use of information and data. In addition, it looks at current security structures and recognizes several critical points where data privacy must be enhanced to reap the long-term benefits of innovation.

**Research purpose**

Many of the problems that the transportation industry faces can be solved by utilising Big Data technologies to collect information about traffic throughout the city and provide realistic data and solutions for traffic management and urban planning, such as congested traffic forecasting, traffic flow prediction, and technical infrastructure planning. Data collection using IoT technology, real-time and historical data collection using significant data technology, data cleaning and preprocessing using extensive data technology, and traffic forecasting using appropriate technologies are all possible.

**Research result**

This paper’s research methods are based on a systematic investigation and evaluation of recent big data breakthroughs and the retrieval of relevant data from literary sources. Different aspects of information gathering are covered by these methods, the first of which are books and articles, research papers, and scientific journals in this field. Another option is to look for publicly available information and other online resources. Identifying research, selecting relevant studies, analysing content quality, extracting data, and synthesising data are all part of the process. A thorough review of the current literature was required to expand the knowledge base on critical issues, create a compelling narrative, and explain the paper’s need, research targets, and direction.

**Summary of the work**

CHAPTER 1. INTRODUCTION TO BIG DATA AND TRANSPORT

This chapter considers the latest developments in Information Technologies and the increase in the data storage means available to us. If we look at the recent technological advancement, it is not unexpected for this type of advancement to lead to massive and increasingly growing amounts of sourced data. This enormous amount of information being generated and collected each moment, along with the challenges and creation of new qualitative data analysis methods, is accepted as a new concept referred to as Big Data.

The logistics and transport industries are better positioned to benefit from Big Data technologies' systematic breakthroughs and analytic tools. Today, as the digitalisation of these sectors progresses, transportation and logistics providers are constantly creating huge and huge datasets while managing the massive flow of goods and people. In transportation, the increase in data is especially evident in the availability of traffic information through navigation applications. A similar passenger information application provides departure information to users of public transport. In addition, payment of shipping costs (ticketing and tolls) is increasingly dependent on data-dependent technologies, applications, and services.

CHAPTER 2. BIG DATA FEATURES, SCALES AND DATA SOURCES

Big data refers to the more diverse information emerging in larger volumes and moving faster. Traditional data processing software is incapable of handling these massive data sets. Companies, including those in the logistics and transportation industries, must use specialised tools known as big data analytics to obtain the best data analysis. These tools enable the management and analysis of massive amounts of data generated by sensors on roads and vehicles, GPS devices, customer applications, websites, etc. Big data analytics’ impact is not limited to the logistics and transportation sectors; other forms of transportation, such as airlines, airports, freight,  and railroads, can benefit from its impact, which will assist in strategic planning, strategic efficiency, and brand management, and customer relationship services.

CHAPTER 3. ITS AND RELATED APPLICATIONS

For operators to improve performance, efficiency, service provision, safety, and security across all modes of transportation, a significant amount and diversity of data is now available. Operators can also use data to manage demand fluctuations, customer support, environmental impacts, and innovation. Traffic signal coordination, trains reporting track defects, online flight check-ins, and cargo tracking are examples of this.

**Scope of work**

The publication consists of fifty-three pages, seven pictures and two tables. Thirty-six works of literature were used in the study. An introduction, three chapters, and a conclusion comprise the dissertation structure.

# CHAPTER 1. INTRODUCTION TO BIG DATA AND TRANSPORT

## 1.1. Literature Review

1. ‘Cognitive Computing for Big Data Systems Over IoT’ by Arun Kumar Sangaiah, Arunkumar Thangavelu, Venkatesan Meenakshi Sundaram

One of the new and developing technologies of the last decade was the Internet of Things. Furthermore, it is one of the fastest technologies to increase in size and quantity throughout the world, as it is already estimated that there are more than twenty billion IoT devices in the world. This book primarily discusses Big Data and IoT alternatives, shining a light on both and discussing their respective research areas. Furthermore, it discusses particular domain expertise, data science techniques, and different IoT advancements.

This research paper delves into the fundamental concepts that reinforce deep learning methods, demonstrating how big data becomes increasingly important as IoT problems evolve. It also encompasses in-depth essential research contributions in data science from both a methodological and an application standpoint, resulting in long-term solutions.

Companies in the logistics and transportation industries must use specialised tools known as big data analytics to obtain the best data analysis. This book is informative and scientific and one of the best resources for learning about current developments in Big Data and IoT technologies.

1. ‘A deep learning approach for detecting traffic accidents from social media data’ by Zhang Z. He Q. Gao J

People are killed and injured in road accidents caused by various factors. These factors can be accidents themselves, bleeding, burning, or not being able to call for an emergency when needed. Sometimes you may call for an emergency, but the said help may not be fast enough to reach and help the unfortunate accident victims because of traffic or road state. Checking the data available to us, it is concluded that most of the casualties resulting from road accidents happen within the next few hours. It means that the most critical moment for the victim is during that time frame, so doing whatever it takes to lessen the amount of response time can lead to life-changing results for many people. Furthermore, if a victim is rendered permanently disabled in a disaster and there is no one nearby to assist, an automatic accident reporting method can potentially reduce the number of such fatalities.

Different methods have been used throughout the years to prevent accidents from happening through the traffic. However, even though many things have been tried, and most of the solutions have been positive, because of the rise in the population and transportation, the number of accidents starts to increase more and more as time passes.

1. “Big data for transportation and mobility: recent advances, trends and challenges” by Ana Isabel Torre-Bastida, Javier Del Ser, Ibai Laña, Maitena Ilardia, Miren Nekane Bilbao, Sergio Campos-Cordobés

Big Data is a new paradigm that has sparked much interest worldwide, particularly in the transportation industry. Emerging technologies and new concepts such as Smart Cities improve the transportation data life cycle. In this context, Big Data is viewed as the transportation industry’s new commitment to effectively manage all data required to provide secure, cleaner, and more efficient transportation options while also allowing users to personalise their travel experience. However, because of the numerous and heterogeneous transportation/mobility application scenarios, Big Data comes with technological challenges. This survey looks at the most recent research efforts in the transportation and mobility industry, including Big Data applications, baseline scenarios, fields, and use cases like routing, planning, infrastructure monitoring, and network design. This analysis will be conducted solely from the standpoint of Big Data, with a particular emphasis on the significant contribution of strategies, tools, and methods for modelling, processing, analysing and visualising Big Data in transportation and mobility. The literature review is used to extract a set of trends and challenges that will provide scientists with a different view on transportation and mobility.

## 1.2. Transportation Industry Challenges & Opportunities

Space, air, land (rail and road), cable, pipeline, and water are all modes of transportation (Figure 1.1). The field is divided into vehicles, infrastructure, and operations. Transportation allows people to trade with one another, which is necessary for developing civilisations.

Transportation infrastructure consists of fixed facilities that we can see and access through the city and towns. They are railways used for long-range transportation, roads and motorways used for vehicles, airways and airports for flights between countries or even continents, bus stations for public transportation, etc. In addition, terminals can be used for both passenger and cargo exchange and at the same time for maintenance. Transport means any mode of transportation used to transport passengers and goods.

The way we plan, design, and construct transportation systems are changing thanks to new technologies. For example, transportation and logistics organizations and businesses use new and emerging technologies to detect accidents, handle transport operations, alter schedules and road plans, calculate the distance and time used for travel, assess the risks involved in the routes mentioned above, etc. In addition, traffic reports, digital maps, real-time traffic situations, delivery apps, take outs, GPS technology, and a variety of other services that did not exist a generation ago are relied upon by travellers.

As seen from these examples, technologies used in transportation and logistics are mostly new technologies or devices, using advanced algorithms and techniques to use them to their full potential, and at the same time gathering way more data and information that is a lot more encompassing the problematic situation. Therefore, these devices, technologies and inventions, including Big Data, Intelligent Transportation Systems, intelligent cities, electric vehicles, and modern sensors, are at the heart of many studies.

Let us imagine a scenario where multiple organizations compete for dominance in the new technology. Of course, most of them can work independently globally, as there are multiple countries and instances where the same product is handled. Nevertheless, regarding the transportation industry, businesses need to determine if the market they intend to compete for is big enough for them, or even if it is big enough, do they have the capacity and luck to go head-to-head with other giants of the industry. However, because we are talking about technology and software development, even the smallest of companies or just one single person can accomplish big things in this part of the market. Moreover, organisations need to consider the political and socio-ecological implications of the said technologies and devices, as every new development can change how we interact with nature, the environment and even each other. Sometimes, new rules and regulations must be created, or old ones must be updated to include these new cases.

Planners and engineers must understand the potential of new technology and its limitations to incorporate it into new project proposals effectively. Furthermore, because technological advances have the potential to disrupt established ways of doing things, technological advancement may need to be accompanied by organisational assessments that enable leaders to remove barriers and support innovation. Studies on various issues can guide these decision-making procedures while also advancing technology.

The progress and technological development have led to significant improvements for transportation companies in the different phases of the delivery process [3]. Logistics, a critical component for transportation companies, has evolved due to technological advancements. Advances in the computational power of new devices and the computer, invention of new methods and techniques regarding data collection and analysing, and usage of more powerful software programs have created the conditions for better and modern logistics and transportation businesses to emerge in recent years. Cars, trucks, and various other vehicles have been armed with new communication and emergency devices, which they can either use for information sharing between different drivers and users or directly with the organization and government facilities it serves. It is now more straightforward for a business to manage the transportation of truck fleets and their personnel. These enhancements enable businesses to comply with prescribed standards, knowledge of pricing, inventory, and procurement, allocating commands between different trucks to optimise deliveries and establishing shorter routes to improve drivers’ daily lives, among other things. That is to say, technological breakthroughs in transportation services have resulted in a plethora of logistical advantages that are beneficial in offering consistent services.

Vehicles, such as cars, buses, and motorcycles, have benefited from technological advancements. This impact has resulted in more excellent dependability for businesses in terms of their truck fleets. Furthermore, onboard electronics offer engineers much information about the health and condition of their vehicles, significantly lowering the risk of failure or breakdown. These new improvements mean increased safety for truck drivers on each of their deliveries. Technology’s contribution to transportation has also positively impacted the environment. Two examples of technological advancements are introducing hybrid technology and the ability to turn engines off when the car is stationary. As a result, trucks can make more secure deliveries thanks to new technology developed daily, using fewer natural resources on our planet. In addition, we have seen a significant improvement in logistics due to the numerous changes brought about by new technology [4]. Logistics is the general process of storing, transporting, and using resources and goods. Identifying potential distributors and suppliers, organizations and businesses, devices and technologies required for these transportations, and calculating all of the benefits and drawbacks of these goods are all part of logistics management [5].

## 1.3. The Impact of the IoT, Big Data and AI

In the current situation, humans are developing numerous technologies to understand our daily activities better to improve our lifestyles by improving our operations. Table 1.2 (List of emerging technologies in Transport [7]) shows many of the technologies developed in this area of expertise. AI, IoT, and Big Data have all contributed to some exciting and intriguing advancements in the transportation business.

Big Data, the Internet of Things (IoT), and Artificial Intelligence (AI) transform how we live. The following technologies are causing waves in industries ranging from transportation to our houses. Moreover, incorporating these technologies into new mobile app development is critical for generating cash. As a result, its benefits and prospects appear limitless.

The IoT is a system of connected physical things that anyone can access via the internet. IoT is a fundamental concept that connects all objects in a person’s environment to the internet. For example, an IoT ‘thing’ could be a human wearing a heart monitor or a car with built-in sensors, i.e.

Because of the creation of new IoT devices, we can use them in the transportation and logistics industry. IoT devices are primarily connected networks of different technologies, so communication and sharing of available data between them will help acquire and transmit data. As there are thousands of different IoT devices, we can show some examples for them that are being used in the transport sector right now:

* Reservation systems
* Toll & ticketing systems
* Security and surveillance systems
* Remote observance

Artificial intelligence (AI) refers to the ability of a computer program or machine to think and learn. The AI concept is predicated on creating robots that can think, act, and learn in the same way that people can.

AI applications in the transportation industry are driving the evolution of the next generation of Intelligent Transportation Systems. Meanwhile, some of the most prevalent aspects of transportation that we employ are (Figure 1. 3):

* Traffic Management Solutions
* Smartphone Apps
* Passenger Transportation
* Law Enforcement

AI has made everything from road safety issues to fleet management system monitoring possible in the transportation business. Nevertheless, Artificial Intelligence (AI) has served as a one-stop shop for various problems.

**Big Data in Transportation Industry.** To begin, Big Data is a term that refers to a massive collection of extensive data. It continues to increase substantially. Second, it is available in three different varieties:

* Structured
* Unstructured
* Semi-structured

As a result, big data can extract information from that data and use it in various sophisticated analytics applications. Big data has changed everything in the last few years. Even the transportation business is helping to make regular commutes easier. On the other hand, Big Data is a rapidly expanding paradigm that has piqued everyone’s curiosity worldwide. Particularly in the transportation sector, such as airlines, trains, hospitality, and other travel and transportation industry aspects.

## 1.4. The Importance of Big Data Analytics in Transportation and Logistics

Big Data is everything. It is a new term but an established concept. It collects, stores and then mines (analysing for patterns) a massive amount of data for competitive advantage. It can be very revealing, and the trail has been blazed by retail giants over the last 20 years. Got a loyalty card anywhere? They are collecting your purchasing data. When you look at massive amounts of it, it throws out some interesting information that can be used to sell more products. These processes are the same in every field. However, looking at the transport sector, we can find some new or odd patterns among the information we collect. For example, we can analyse accidents happening during a specific day, such as during rush hours on 1 of the major roads. With this information at hand, we can change or alter the roads or make them safer for driving. It would help if you had tons and tons of data before these patterns become apparent, and it is now straightforward (from a technical standpoint) to collect and farm this amount of stuff, hence the Big Data trend. In the recent past, you would have needed lots of costly hardware and intelligent people to run it, so it would only be available to the giants like Walmart.

Data capture all of the seemingly trivial facts of modern life. Data evolves and expands with each passing day, from studying your bank account spending habits to more extensive, more complex processing capabilities. Considering how intimately people engage with their digital selves, you could easily mistake humankind for a hybrid of digital and biological stuff. This is not a fantasy discussion, either. It is accurate, and the amount of data produced every 48 hours rivals the total amount of data and information accumulated over thousands of years of human history, as Bernard Marr explains [8]. Big data in the supply chain has become associated with more excellent business, supply chain efficiency, continuous improvement, and innovation.

Big data has been trending during the last few decades because it is accessible with cheap technology. Unlike mid-nineteenth-century computer systems, modern data processing is more cost-effective than calculating data on paper. Furthermore, keeping data digitally is less expensive than storing physical copies of data, especially when climate control, security, and accessibility are considered. According to Forbes magazine’s Lisa Arthur, big data can take numerous forms, including unstructured and multi-structured data. These many forms significantly impact what and who creates and uses big data.

Before you consider future possibilities of Big Data in the transportation sector, you must first comprehend how it is currently utilised. As a result, supply chain organisations will be able to reduce total operational expenses, develop and test new business models, build and sustain a higher level of customer service, and reach unrivalled efficiency, according to Katrin Zeiler [10]. Figure 1.5 depicts the full spectrum of big data’s impact on the transportation, supply chain, and logistics industries [11]. The utilisation and usage of Big Data technologies in this sector make it critical to understand what implementing said technologies mean for this industry. Not only for the transport sector but also in logistics and supply chain.

1. **Operational Capacity Planning:** short- and medium-term capacity planning enables practical resource usage and growth.
2. **Crowd-sourced pickup and delivery:** Many scattered carriers pick up and deliver packages along their regular routes.
3. **Service Improvement:** Customers' and users' experience, information and data are collected through the operation, and b getting feedback from others, we can generate special requirements and further improve our customer services.
4. **Consumer Loyalty Management:** To predict attrition and take countermeasures, public customer information is mapped against business characteristics.
5. **Risk Assessment and Resilience Planning:** The resilience of transportation services is improved by tracking and predicting events that contribute to supply chain interruptions.
6. **Intelligence on the Environment:** We gather information from devices and sensors put on the vehicles by organizations. These collected data help us generate real-time geoinformation regarding the environment, and later with different data, we can fine-tune it to increase its accuracy further. Later this information can be used by either the identical vehicles or for future references for the area.
7. **Address Verification:** Customers' and users' address is essential and sensitive information. At the same time, organizations need correct info on address in order to be able to deliver goods or deliveries to the correct location.

In the transportation industry, big data decreases delivery and pickup errors. Industry professionals agree that courier service firms are often major operational organisations dealing with enormous cargo volumes, hub terminals, general information systems, and various vehicles. Furthermore, they are comprised of a complicated web of personnel and equipment. Errors in deliveries, pickups, and shipping operations, on the other hand, can result in increased costs for the organisation. While these costs may seem modest, consider how much money is lost every day if a single error occurs. This may cost more than necessary, and customers may request replacements or refunds, causing the company’s reputation to suffer. However, delivery and pickup errors can be eliminated with big data. Therefore, sensors are integrated into all delivery vehicles used by logistics companies, with GPS-enabled cellphones filling up the gaps.

When logistics businesses are competing for new contracts, the dependability and timeliness data from these sensors is used by a third party to check them for correctness. So instead of causing a corporation to lose money due to errors, big data may assist the organisation in increasing profitability.

The use of big data will boost operational efficiency. One of the most known benefits of Big Data in transportation is increased efficiency and optimization of resources and data. In most cases, automated data processing enhances decision-making capabilities, process quality and performance, and resource consumption. The advantages can be extended in the following ways.

* Ability to track and forecast low-in-stock items ahead of time
* Reduce the impact of late and partial shipments significantly.
* Examine the potential for a variety of events to emerge.

Big data enables improved shipping alternatives with higher product quality. However, when it comes to shipping sensitive items, the best route must be chosen first to avoid late delivery and ensure that the goods remain in good condition when they arrive at the consumer. Drivers and organizations need to determine the best approach to the route they will take. Different variables affect this information, such as weather, location, time, the season of the year, and temperature. At the same time, if we can get information from other users who have already travelled this route, we can better determine the condition of the road and travel. If we can gather all of this data, we can analyse it and come to a different conclusion, which will help determine the most cost-effective route to arrive at our destination.

Indeed, big data combined with the Internet of Things might provide delivery drivers and managers with a much better understanding of how to avoid expenditures associated with perished items. For example, installing a digital thermometer within the truck to carefully inspect the items inside and sending the data to a central routing workstation and traffic data is an option. The computer might then warn the driver if the driver’s intended path would harm the fruit and vegetable and recommend different routes instead. When you consider these advantages, you will realise that they derive from three main areas in the transportation business that are touched by big data, as shown in Figure 1.6.

As seen from the above information, we can see the impact of Big Data in the logistics and transportation industry. Furthermore, it takes a big part of modern and future technologies in these fields of study. By studying and inventing new techniques and methods, we can define new options for different conditions. Big Data in transportation is not simply collecting and storing data, but it uses it anywhere it can use, sometimes even being able to get the data that has not been collected but is being correlated and calculated by other related data. Furthermore, as in every industry, organizations and businesses mainly need to focus on the users and customers to make it easier for everyday people to use these technologies and focus on the customer aspects of the development.

# CHAPTER 2. BIG DATA FEATURES, SCALES AND DATA SOURCES

## 2.1. Big Data’s 3V

When discussing Big Data, three main characteristics are mentioned: volume, velocity, and variety, derived from Laney’s 3 V model [12]. These elements are critical in how data is ingested, analysed, managed, stored, distributed, and processed. The following list and Figure 2.1 describes in detail the three most important indicators that distinguish Big Data from “normal” data [13]:

1. **Volume:** The total amount of data captured, managed, and analysed is volume. Nowadays, it is not uncommon to find experimental setups dealing with volumes up to the yottabyte scale (Y.B., 1024 bytes), with petabytes (P.B., 1015 bytes) as the minimum data volume for a scenario to be recognised as Big Data.
2. **Velocity:** this feature can refer to two concepts: (1) the rate at which data instances are generated and (2) the rate at which such data samples are received and processed. Currently, investigated ITS scenarios necessitate real-time processing latencies in nanoseconds, a stringent constraint requiring critical design decisions in the computing infrastructure and the developed data analysis models.
3. **Variety:** the heterogeneity of the captured data is due to the diversity of digitalised domains related to the application or service at hand. The requirement for jointly processing and analysing them has significant implications for system design, including data integration and fusion functionalities.

Although the above description of Big Data features is not limited to any particular field of interest, as seen in Kwan M. P’s study [14], it is a good starting point for narrowing them down to the domains addressed in this paper: transportation big data.

In terms of volume and size, Big Data refers to datasets with sizes that make traditional approaches to collection, management, processing, and analysis impractical. Over time, datasets have grown in size, with single datasets varying from a few terabytes to man petabytes. When it comes to transportation and mobility, an excellent example of new geo-localised data sources with increasing relevance in these fields is social media. For example, Twitter has announced that its users send over 500 million tweets per day, whereas Instagram claims that its users shared about 60 million photos per day in 2014. Some of this media content is geotagged with GPS data, allowing us to infer people’s behavioural trends over time.

Similarly, the volume and coverage of sensed transportation data are growing due to the widespread use of sensors in vehicles, wearables, cell phones, and other devices. Sensors are deployed on roadways in major metropolitan areas that were not previously instrumented or even visually inspected [15]. In addition, smaller communities with limited or no data collection are gradually deploying sensors and implementing data collection mechanisms, such as crowdsourcing. Overall, the increased number of sensors results in a significant increase in data volume in all of these cases.

Big Data technology enables data analysis while it is being generated rather than after it has been stored in databases. Real-time data stream processing for traffic control is one example. In terms of data variety, many opportunities now arise from capturing massive amounts of data from various sources and the ability to connect and exploit them with people, governments, and businesses. Traditional legacy systems, relational databases, sensor networks, Open Data, email, video, blogs, call centre conversations, and social media are just a few examples. Because of the diversity of data sources, the type and form in which data are represented vary greatly; indeed, even for a fixed nature of the data (e.g., high-definition radar signals for vehicle speed and count detection), differences between the data delivered by sensors capturing them may appear simply because different companies manufacture them.

Big Data represents a significant increase in the quantity, variety, and accessibility of data. There are three primary Big Data sources in transportation and mobility: unstructured social media, sensor data (unstructured or semi-structured), and open data (unstructured – raw text, semi-structured – JSON/XML, or semantically structured – Linked Open Data).

1. **Social Media:** Nowadays, Social Media platforms store enormous amounts of data that could reveal critical information upon their eventual analysis. The underlying reason is that the user role has shifted from being a mere consumer to providing content assets. Social media is a well-known technology in our modern world and population, as its one of the most used technologies for communication, connection, and information sharing. Social media is a web-based application created for users to share and access other users' posts. Therefore, social media can be considered a context-rich source of Big Data, usually referred to as Social Big Data. Furthermore, it is an important data source in transportation and mobility because the gathered information is typically geotagged, allowing for the inference of geo-localised knowledge about the mobility of the people who produce data traces. Other data sources in this category include collaborative applications and initiatives such as crowdsourcing, which is the practice of gathering information or input for a task or project by enlisting the help of a large number of people. Logistics is another area where this type of data is widely used nowadays, giving rise to novel concepts and collaborative services. Among them, crowdsourcing delivery refers to the scenario in which citizens deliver goods to each other along their route to reduce the carbon footprint of the delivered product while keeping in mind that deliveries must be done with a minimal detour from the deliverer’s original path [16].
2. **In terms of sensor data**, the proliferation of physical devices embedded with electronics, software, sensors, and actuators fueled by the IoT paradigm is dramatically accelerating the pace of innovation in the transportation industry, particularly in concepts such as the connected car (also known as automobile as a sensor), which allows for the exchange of sensed information not only between vehicles but also between vehicles and infrastructure. Furthermore, new modes of last-mile delivery using new connected means, such as drones or autonomous vehicles, are expected to benefit supply chains and logistics operations significantly. Because of the criticality of an accident in an automotive scenario, transportation safety is another topic of increasing relevance for the IoT domain. To summarise, IoT is expected to be a critical technology in current and future transportation and mobility management solutions, as it can improve safety and efficiency, reduce environmental impact, and foster new services and capabilities. Similar claims have been made in research related to urban computing, which refers to the acquisition and integration of data generated by various sources in urban areas. Ubiquitous sensing technologies (e.g., IoT) play an essential role in urban computing, mainly when dealing with transportation issues throughout the city.
3. **Open Data:** Both citizens and the public sector benefit from open transportation data. Therefore, governments should publish open transportation information in machine-readable and easily consumable formats. Many countries have recently embraced this potential by creating data repositories for national public transportation services rather than transit systems.

## 2.2. Data Collection

In today’s society, humans are using modern equipment and devices worldwide. Moreover, all these technologies collect and store their usage and user information, which is sometimes useless. However, at other times these pieces of information are vital and concern the security and privacy of customers themselves. This means that, even if a person does not want to leave behind any data regarding themselves, because of the modern technologies, there will still be data generated regarding them, such as them leaving behind footprints whether they want it or not. After this information is collected, sometimes they are discarded immediately and considered useless, and other times, they are sent to the central databases to be stored, then later used and analysed. Considering the amount of information generated each second by one person, and later by combining this with other relevant data such as geolocation, time and space of the data generated, we can accomplish many different results and come to different conclusions [17].

There are mainly two methods of generating data: digital and analogue data. They are differenced by their use cases and purposes for the generation. Digital data is created to mainly be used by computers and machines, devices with software that can specially create, store and analyse the generated data:

* Commercial transaction data
* GPS (Global Positioning System) or other geo-located spatial data stamps.
* Data generated by vehicles, cars, electric vehicles, different transportation devices and so on
* SMS and Emails
* Cards and other tickets are used for transportation and logistics purposes

Digital data is created with specific aims in mind, and they will be used for meeting different criteria or specific needs. In the last century, because of the costs of data storage and computational power, we tended to store the bare minimum of information that is useful to us, such that any other information, even though it might be used for other purposes, was not collected and stored. Nevertheless, with the advancements in data storage, cloud technologies, and computational power available to us, nowadays, simply storing and collecting data is not a problem. This entails that we do not need to worry as much as before, and we can collect a lot more information and use them to get better and more accurate information.

When physical phenomena, such as sound, light, vibration, etc., are related, we can later store this information. We call analogue data and collect them using different sensors and technological devices. Microphones, cameras, special health devices, and so on can be shown as an example. With the increase in the production and quality of sensors, analog data is expanding rapidly, including all other fields. They can be used in everyday tasks, scientific experiments, human health problems, and even collecting information about deep space.

Nevertheless, analogue data is not omnipotent at the same time, as its limitations still restrict it. One of them is the devices they are collected by, sensors. Sensors will have their purposes and use cases, such that a sensor for recording sound will not be recording light or some magnetic fields, and vice versa. This means that the data we collect will depend on our sensor. Some of the sensors can be combined to record multiple sources of information, but they are still not all-encompassing, but only the combination of multiple sensors. However, with recent studies and advancements in science, it has been proven that some information can be deducted from other sources of data. This means that, as long as we can find methods for generating and calculating other types of information, we may be able to collect multiple data with only one source of information.

The second limitation of analogue data is their storage, longevity and transmission, such that, even if we can collect massive amounts of data, if we are not storing and analysing them, they will still be useless for us later. An example of this can be shown by engines used by different vehicles, as they can generate terabytes of data per hour, but after using this data and getting results, they will be discarded immediately, as the next wave of data needs to be stored and calculated again. Therefore, we need new algorithms for compression and storage purposes, as its one of the vital parts of analogue data.

Filtering and compression-related data loss are becoming less an issue as low-cost data storage and transmission technologies combine with cutting-edge database platforms like Hadoop. These developments, coupled with the increasing complexity and lower cost of dual sensor-computing devices, increase the significance of stored data.

Big Data extracts social or economic value from large-scale, diverse, and complex data streams using various technologies and system architectures. Four main technological developments are at the heart of them:

1. **Sensor platforms**

The recent increase in data generation is thanks to the millions of new devices and vehicles being equipped with new generation sensors and transferring the generated data to the datacenters without deleting the acquired data.

1. **Real-time data analysis**

With the advent of new algorithms, increased storage capacity, and real-time data generation and analyses, we can inquire about intended information with minimal time loss equal to real-time analyses. This helps a lot in the time-sensitive fields and transportation sector, as sometimes even a millisecond can determine between safe travel and a horrible accident.

1. **Advances in data storage**

Data storage capacity has been increasing at an increasing rate through the years, so much that it would be unimaginable before. We can continuously generate and store petabytes of information, increasing by the second. One of the biggest challenges faced by the industry was how to store such vast amounts of data, and because of these problems, there would be much digital dust that would be generated and later wiped out without seeing any use. Nevertheless, now, we can store any amount of information we want, and at the same time, because of advanced technologies and software programs, we not only store but acquire, query and handle the said data with ease. So what this means is that, with new methods and techniques, we can acquire new kinds of information from digital dust, as they have never been used before, and they can give us some surprises, making us see new angles and implementations for the generated data. Talking about these things can lead to one big question: where are all of these data stored and secured? Even though we have advanced storage technologies, there still needs to be hardware that stores all of this information. That question is answered worldwide, in different countries and data centres. With the emergence of cloud technology, even without having personal data centres, we can use the vast amounts of data centres in the world as our own. Another benefit of this is that you do not need to attend data centres or be responsible for their security, as they are all being handled by the data centre themselves.

1. **New analytic frameworks**

One of the problems technology faced before was not being able to perform large amounts of work due to inefficient and low rate of computational power. In order to combat this, other than having new and robust hardware, new and more advanced software programs with optimized techniques, methods, and algorithms need to be created. With the newly invented algorithms, searching and analysing times for Big Data can be optimized so that it is simply unbelievable. We can show Hadoop and other similar technologies as examples of this, as they are using parallel processing, taking apart significant works, separating and working with each of them simultaneously, using different computers. This helps with optimization and performance while reducing work's running time. They later distribute these processed tasks to other connected devices, such as with more and more devices connected to the network, the better and faster it will become. These new algorithms and techniques can be applied to any data, regardless of the amount and type of data set. The only downside of these techniques may be their inability to be used in real-time processing, as it takes much time for the results to be considered real-time.

## 2.3. Real-time data analytics

Existing Intelligent Transformation System infrastructures are not intended for real-time data processing, nor are they capable of analysing collected information at the rates required by critical systems. This situation creates a problematic paradox because, for most transportation problems, a fast response time is vital to maintain that information and decision making are closely linked in time and thus valuable for practice. Three broad categories of Big Data systems can be identified based on the required latency level [18]:

* **In Batch systems**, the data processing cycle is subject to very loose (hours to days) or no latency constraints. These systems collect data, process it, and generate results in stages. This is the case of delay-insensitive applications, which are exemplified in the context of ITS by control panels for medium-term (e.g., weekly) freight transportation monitoring and planning.
* **Soft real-time systems** have weak real-time constraints ranging from seconds to minutes. Online systems are also included in this category, as the response time to the user should be quick but not critical. A query service on the occupation of parking lots in a Smart City is an example. While a more responsive user interface with the system improves the user experience, compliance with real-time latency constraints is not required for the application at hand.
* **Hard real-time systems**, in which meeting ultra-low-latency constraints is a hard constraint, with thresholds ranging from milliseconds to nanoseconds. If the designed data processing life cycle does not meet this timing requirement, the developed system fails and cannot support the application for which it was designed. Examples of latency-critical applications include incident prediction and inference of the state of the road for autonomous vehicles

As a result, real-time processing is regarded as critical for enabling new forms of on-demand transportation for people, necessitating real-time data sharing and interpretation, and necessitating legislative and public opinion shifts to implement intelligent mobility use cases utilising real-time network capacity data for people, vehicles, users, and so on. Service automation and optimisation in real-time will be critical attributes for future intelligent transportation solutions. The gap to be bridged is not the technological transformation of this sector but rather the transformation of people toward processes and methodologies for using and profiting from future ITS. Real-time systems must be quick. If they are not fast enough to process data as it comes in, the system will be overloaded, and the system will no longer work in real-time.

There can never be a natural real-time system as even the fastest transmission of information between the components of one device will take time, even if it is negligible in the grand scheme of things. Moreover, if we were to multiply this by hundreds of thousands of devices and kilometres of distance between them and data centres, we can already imagine that the time lost there will be at least milliseconds. So what this means is that no system in this world can be real-time. Nevertheless, what we mean by saying real-time systems is that the delay is so slight that it is not considered and is within the acceptable range of the procedure happening at the moment. For example, a bank ATM’s data processing can be considered real-time if it responds in less than one-tenth of a second. In a supercomputer, however, the same speed would be considered slow. Real-time processing is used when you need the output immediately, and it has advantages and disadvantages.

Advantages of Real-Time Processing:

* The information is continually updated and ready to use.
* You have increased uptime.
* The delay in data processing is minimal.
* System synchronisation would require fewer resources.
* It aids in identifying problems, allowing you to take immediate action.

Disadvantages of Real-Time Processing:

* It adds a data overload in the event of a system failure.
* It requires high-performance hardware and is expensive.

## 2.4. Security and privacy

Security and privacy are long-standing and widely researched issues in transportation. Because of the personal geo-localised data and content shared by users via their devices and computers, privacy is likely the most stringent security requirement imposed by transportation applications leveraging Big Data. As a result, the transportation industry continues to face unresolved challenges in terms of collecting private data. Security flaws and gaps range from identity theft to unwanted location and data tracking. Furthermore, trends such as using payment data as a secondary source to predict macroscopic mobility patterns massively increase the associated confidentiality risks with the Big Data framework. These privacy concerns jeopardise the widespread adoption of Big Data technologies for ITS and mobility; surprisingly, we find a scarcity of references and contributions with effective ITS technologies that ensure confidentiality, integrity, and secure data handling, including financial and personal information. Without a doubt, tackling this challenge with specific, viable technical approaches would cast a thick blanket of certainty over the use of Big Data in transportation and mobility, assuring users that their rights are protected following international laws, directives, and recommendations.

Customer trust is also affected by big data privacy. The more information you accumulate about customers, the simpler it is to “show the correlation” and understand their behaviour and actions, predict future actions, and build deep and detailed profiles of their lives and preferences. For this reason, it is essential to be open and honest with your customers about what you are doing with their information, how you will store it, and what steps you are taking to comply with data protection regulations.

The architecture of data collection methodologies and procedures is central to the most debate over the collection and use of personal data. Most of the time, little is expected to change how data and data systems effectively implement and transfer bits of personal information. However, this viewpoint is challenged, particularly by proponents of the “Privacy by Design” approach. According to this point of view, data collection systems and procedures must be designed from the ground up also to include solid and irreversible privacy safeguards for data collection systems.

“Privacy by Design” was developed to design data collection mechanisms and practices. “Privacy by Design” simply means “data protection through technological design.” The idea behind this is that sensitive data in the data processing stages is best applied when it is already built into the technology when it is created. Nonetheless, there is still confusion about what “Privacy by Design” entails and how it can be implemented.

# CHAPTER 3. ITS AND RELATED APPLICATIONS

## 3.1. Intelligent transportation systems in big data

Intelligent Transportation Systems (ITS) are supervision and information systems that use incorporated communication systems and data analysis technologies to improve the mobility of people and goods. Enhancing safety, reducing traffic congestion, and effectively managing incidents. The evolution of ITS is still ongoing. Thereby, the extent to which these technologies are used – and the sophistication they are deployed – varies by country. Intelligent Transportation Systems (ITS) are information and control systems that use data processing and integrated communications technologies to accomplish the following tasks:

* enhancing people and goods mobility
* improving safety, reducing traffic congestion, and effectively managing incidents
* achieving transportation policy objectives and goals, including demand management and public transport priority measures

The definition encompasses a wide range of techniques and approaches that can be accomplished through stand-alone technological applications or integrating various systems to provide new transportation services. ITS is critical in the frame of reference of road network operations because it provides tools to improve safety and transform mobility. Wireless, digital, and automated technologies are all part of ITS. These technologies, when combined, have the potential to include automobiles, users, and infrastructure. Many ITS technologies can help with trip optimisation by reducing unnecessary miles travelled, increasing other modes of transportation, reducing the amount of time expended in traffic, minimising dependence on foreign oil, and cleaning the air. Moreover, once applied to the management system and car design, ITS technologies have the potential to decrease the consumption of fuel by:

* allowing for optimised route planning;
* increasing the appeal of using public transportation;
* smoothing accelerations/decelerations and stop-and-go driving;
* reducing congestion;
* allowing for small platoons of closely spaced vehicles (Weight reduction may be possible with safer vehicles without jeopardising occupant safety);
* Modifying vehicle transmission to account for changing road conditions and terrain;

Many ITS applications play a role in efficient road network operations, with the following goals in mind:

* maximising the capacity of the existing road network
* ensuring that the road network runs as efficiently, safely, and sustainably as possible

The most commonly adopted ITS applications aim to improve road networks’ efficiency, safety, and sustainability. For example:

* systems for managing travel demand and traffic – such as incident management, traffic control, travel demand management, electronic payment, parking management and control
* traveller information system applications that enable road users to make informed travel decisions, such as route guidance and driver information

## 3.2. Applications and related projects

Detailed transportation data can now be accessed and shared at unprecedented scales, revealing new transportation paradigms and opportunities related to smart mobility. The Big Data paradigm fosters many application areas in the transportation and mobility domain, involving some - or all of the critical stakeholders in their application scenarios: citizens (pedestrians and road users), the public sector (governments and authorities), and the private sector. The inclusion of the Big Data paradigm in transportation applications and users does not significantly differ from traditional application scenarios based on data capture, analysis, and exploitation. However, suppose the methodology and technological approaches used in such applications change. In that case, the participation of each stakeholder may span the entire Big Data cycle, from data collection to knowledge inference from the captured data.

After establishing that the data phenomenon is not radically new but relatively incremental in transportation and mobility, we develop a classification of recent Big Data paradigm applications, sorted based on the traditional ITS scenarios to which each of such new applications is related. They are shown in Table 3.1.

**Big Data schemes in social transportation systems:**

Big data for social transportation provides us with unprecedented opportunities to solve transportation problems that traditional approaches are incapable of solving and build next-generation intelligent transportation systems. However, even though social data has been used for transportation analysis, many challenges exist. First, social data evolve and contain a wealth of information, necessitating the collection and cleaning of data. Second, new data mining and processing techniques, such as natural language processing and streaming data analysis, are required to use real-time traffic data [20] effectively.

The expanding use and impact of technology put pedestrians in greater danger than ever before. To assess the impact of interference on crash risk, high-quality data on the pervasiveness of interruption between many drivers and pedestrians are required. In addition, to investigate the role of pedestrian and driver distraction in crash frequency and severity, valid surrogate steps of traffic conflict and state pedestrian crash data are required [21].

## 3.3. Open challenges

The extensive activity evident in the previous sections’ literature review is a clear indicator of the recent technological advances with Big Data applied to the transportation and mobility domains. However, not only do several open research issues remain unresolved and unaddressed, but new challenges emerge as a result of data exchange and exploitation across heterogeneous transportation areas. Several research niches are enumerated and argued about concerning the noted contributions in the last year and postulated what associated problems should be resolved under new Big Data functionalities.

Section 2.2 previously described new data sources emerging in the transportation and mobility landscape that unleash new possibilities, services, and applications, owing to the increased density and heterogeneity of sensors deployed in urban scenarios (e.g. high-definition vehicle speed detectors, pedestrian counting devices and vehicular LIDAR radars for obstacle identification). However, this variety of new data sources brings with it new problems to solve in terms of volume and coverage, variety, value or purpose, and quality, as detailed below:

* Data volume and coverage: today’s transportation datasets are much larger than before. Although specific safety applications do not require high data volumes to be captured (for example, blind-spot warning), other services and applications on board are expected to go beyond the rate limits imposed by current communication standards such as DSRC, such as high-resolution proximity mapping with LIDAR radar sensors. In addition, conventional data techniques and technologies do not work in practice because of the augmented data derived by vehicles, requiring a reconsidering and remodelling of ITS to accommodate such large volumes of captured data.
* Data variety refers to the diversity of data in a data set. It is recognised as a crucial aspect of data complexity. Underlying the variety of transportation data is the challenge of combining, fusing, or integrating it for later analysis or representation. The next step after data fusion is information fusion, which determines how decisions are made based on results from multiple data sources. Optimisation algorithms, for example, can be an efficient option for determining how to fuse and aggregate the captured information streams optimally.
* Data value or purpose: In general, it is not only the amount of data that matters in the Big Data life cycle but also its intelligent use in confirming or validating a hypothesis. This statement is also true for the transportation domain, but with the added disadvantage of a potentially costly data collection process (mainly when physical sensors are deployed along with the infrastructure) [28].
* Data quality is a measurement that evaluates data state using criteria like accuracy, comprehensiveness, consistency, reliability, and punctuality. Unfortunately, quality and accuracy are less controllable in many forms of Big Data, as large data volumes frequently compensate for lack of quality or accuracy. Furthermore, the requirements for accuracy, cleansing, and certainty vary depending on the data-based application at hand: accidents, for example, require as much accuracy and reliability as possible for the processed data, whereas other user services, such as predictive rerouting, do not.

# CONCLUSION

The Big Data conceptual framework will soon become a core key component for ITS. As a result, practitioners and researchers in this field have recently focused on mastering advanced tasks and gaining valuable skills across the Big Data life cycle, particularly data collection and analytics. This manuscript has critically examined the newest breakthroughs in implementing Big Data ideas, technologies, and techniques for transportation and mobility.

Chapter I:

Chapter 1, which is named “Introduction to big data and transport”, talks about the main research area for this paper. This chapter gives information about Transportation, Big Data, and the usage and necessity of Big Data in the Transportation industry. Then, the subchapter “Transportation Industry Challenges & Opportunities” gives information about the problems faced by transportation and logistics and how to handle them. The subchapter that follows, “The IoT, Big Data, and AI”, informs us about new technologies and how they affect the transportation industry. Finally, the following subchapter, “The Importance of Big Data Analytics in Transportation and Logistics”, talks about Big Data and some of its components. It goes over a few key variables and how they affect the supply chain and transportation industries.

Chapter II:

The second chapter is all about Big Data, its features, key components and data sources for Big Data. Each of this chapter’s subchapters explains different aspects of Big Data. This chapter gives information about how such large amounts of data can be acquired, stored and analysed. Furthermore, it talks about real-time analytics and its advantages and disadvantages. The last subchapter, “Security and privacy”, is one the most concerning aspects regarding new technologies, as it directly involves all levels of industry and the one most affecting customers and users alike.

Chapter 3:

The last chapter, “ITS and related applications,” talks about specific Big Data technologies used in the transportation and logistics industry. Intelligent transportation systems are one the technologies related to transport, and they are getting more and more mature as this industry grows. The extensive activity evident in the previous sections’ literature review is a clear indicator of the recent technological advances with Big Data applied to the transportation and mobility domains. However, not only do several open research issues remain unresolved and unaddressed, but new challenges emerge as a result of data exchange and exploitation across heterogeneous transportation areas.

Based on the literature review results, we conclude that the exploitation of new data sources and the development of new Big Data Analytics models to create valuable data-intensive services and applications for transportation and mobility is a critical aspect that deserves more attention in the future. Indeed, the number of companies and startups utilising the Big Data paradigm’s capabilities for the above domains has increased dramatically over the last decade. According to the community, more research should ensure that the Big Data technologies portfolio grows mature, scalable, and functional enough to support applications and services of practical value for new companies and business models. The cornerstone of Big Data is to capture, ingest, and mine massive substrates of evolving, heterogeneous data (as provided by Smart Cities and IoT environments, for example) and thus improve the user experience in transportation and mobility to previously unimaginable levels. If new disruptive Big Data technologies are not designed to ingest and process data with increasingly challenging properties in terms of volume, heterogeneity, and dynamism, they will fail when applied to transportation and mobility problems.

# REFERENCES

1. Transport - <https://en.wikipedia.org/wiki/Transport>
2. Transportation Technologies for the 21st Century by Elizabeth Deakin, Professor in the Department of City and Regional Planning at the University of California, Berkeley
3. Transportation Industry Challenges & Marketing Opportunities In 2022 by Team Linchpin
4. The Benefits of Technology for Transport Services -<https://www.econonord.com/en/2016/10/benefits-technology-transport-services/>
5. Trends Shaping the Future of Logistics In 2022 by Team Linchpin
6. Impact Of Ai, IoT and Big Data on Transportation Industry by SYNARION IT - <https://www.synarionit.com/blog/impact-of-ai-iot-and-big-data-on-transportation-industry/>
7. List of emerging technologies in Transport -<https://en.wikipedia.org/wiki/List_of_emerging_technologies#Transport>
8. What’s Big Data in Supply Chain & Logistics? Why Should I Be Looking at This Tech Trend? - <https://www.globaltranz.com/big-data-in-supply-chain/>
9. What Is Big Data? by Lisa Arthur
10. The Distinct Impact of Big Data Analytics In Transport & Logistics by Daniel Moayanda
11. What Is the Impact of Big Data in the Transportation & Supply Chain Industries? - <https://www.globaltranz.com/big-data-in-the-transportation/>
12. Laney D.: ‘3D data management: controlling data volume, velocity and variety’
13. Demchenko Y. De Laat C. Membrey P.: ‘Defining architecture components of the Big Data ecosystem’. Int. Conf. on Collaboration Technologies and Systems (CTS), Minneapolis, MN, USA, 2014
14. Kwan M.P.: ‘Algorithmic geographies: big data, algorithmic uncertainty, and the production of geographic knowledge.’
15. Collier W.C. Weiland R.J.: ‘Smart cars, smart highways’, IEEE Spectr.,
16. Kaplan A.M. Haenlein M.: 'Users of the world, unite! The challenges and opportunities of social media
17. Big Data and Transport: Understanding and assessing options
18. Real-Time Processing: Difference & (Dis)Advantage Over Batches by SentinelOne - <https://www.sentinelone.com/blog/real-time-processing/>
19. “Privacy by Design” by Ann Cavoukian, Executive Director of the Institute for Privacy and Big Data at Ryerson University, Canada, and former Information and Privacy Commissioner for the Province of Ontario.
20. Zheng X. Chen W. Wang P. et al.: ‘Big Data for social transportation’, IEEE Trans. Intell. Transp. Syst.
21. Methodology for collecting naturalistic observation data of pedestrian and driver interactions - Scopatz RA, Zhou Y, Johnson KL. Transportation Research Board 95th Annual Meeting
22. Banerjee T.P. Das S.: ‘Multi-sensor data fusion using support vector machine for motor fault detection’, Inf. Sci.
23. Hosagrahara A.: ‘Improving engine and vehicle design using Big engineering data analytics and Matlab.’
24. Fridman L. Brown D.E. Angell W. et al.: ‘Automated synchronisation of driving data using vibration and steering events’, Pattern Recognit. Lett.
25. Taie M.A. Moawad E.M. Diab M. et al.: ‘Remote diagnosis, maintenance and prognosis for advanced driver assistance systems using machine learning algorithms’, S.A.E. Int. J. Passenger Cars-Electron. Electr. Syst.
26. El Faouzi N.E. Klein L.A.: ‘Data fusion for ITS: techniques and research needs’, Transp. Res. Procedia
27. Hernández de la Iglesia D. Villarrubia González G. Santana P. et al.: ‘Multi-sensor information fusion for optimizing electric bicycle routes using a swarm intelligence algorithm’
28. Fosso Wamba S. Akter S. Edwards A. et al.: ‘How ‘Big Data’ can make a big impact: Findings from a systematic review and a longitudinal case study, Int. J. Prod. Econ.
29. Big data defined by Oracle - <https://www.oracle.com/big-data/what-is-big-data/>
30. Big data for transport and logistics: A review - 2017 International Conference on Advanced Systems and Electric Technologies
31. Big Data in Transport - <https://www.theiet.org/media/3430/data-trans.pdf>
32. Yang Z. Wu C. Zhou Z. et al.: ‘Mobility increases localizability: a survey on indoor wireless localisation using inertial sensors’, ACM Comput. Surv.
33. Harris I. Wang Y. Wang H.: ‘ICT in multimodal transport and technological trends: unleashing the potential for the future’, Int. J. Prod. Econ.
34. Das A. Dash P. Mishra B.K.: ‘An innovation model for smart traffic management system using internet of things (IoT)’, in Arun Kumar Sangaiah Arun Kumar Thangavelu Venkatesan Meenakshi Sundaram (Eds.): ‘Cognitive Computing for Big Data Systems Over IoT’ (Springer, New York, 2018), pp. 355– 370
35. Zhang Z. He Q. Gao J. et al.: ‘A deep learning approach for detecting traffic accidents from social media data’, Transp. Res. C, Emerg. Technol., 2018, 86, pp. 580– 596
36. Torre-Bastida AI, Del Ser J, Laña I, Ilardia M, Bilbao MN, Campos-Cordobés S (2018) Big data for transportation and mobility: recent advances, trends and challenges. IET Intell Transp Syst 12(8):742–755

# APPENDIX

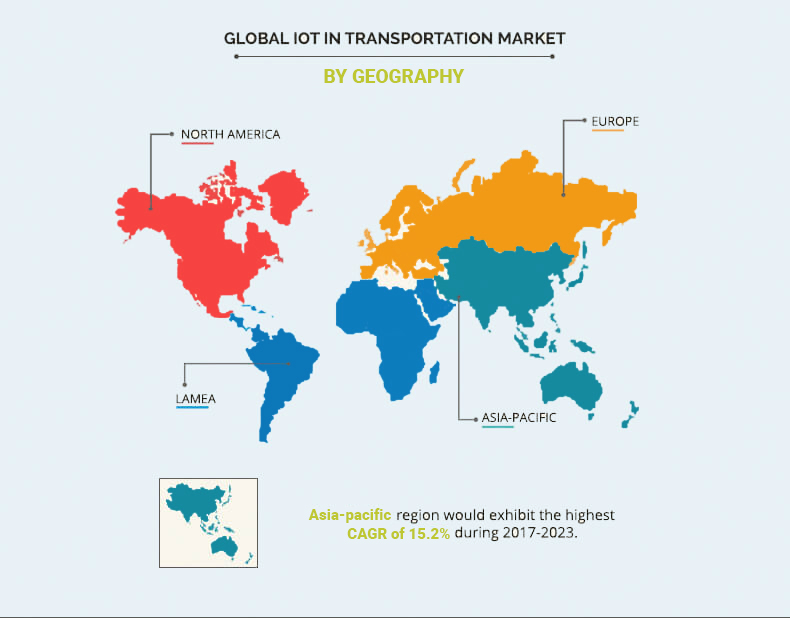
**“All modes of transportation” **

**Figure 1.1****.**

**Table 1.1.**

**List of emerging technologies in Transport**

|  |  |  |  |
| --- | --- | --- | --- |
| **Emerging technology** | **Status** | **Potential applications** | **Related articles** |
| Airless tire | Research, development, early prototypes | Safer tires | Tweel |
| Autonomous Rail Rapid Transit | Commercialisation, diffusion | Reducing air pollution, decreasing oil consumption | Electric vehicle |
| Electro hydrodynamic propulsion | Research, development, prototypes | Better flying transportation, efficient propulsion in air | Electrohydrodynamics |
| Flexible wings (X-53 Active Aeroelastic Wing, Adaptive Compliant Wing), fluidic flight controls | Experiments, prototypes | Controlling aircraft, ships | Aircraft flight control system, BAE Systems Demon, fluidics |
| Flying car | Early commercialisation, prototypes | More effective transportation | Terrafugia Transition, Moller M400 Skycar, Urban Aeronautics X-Hawk, AeroMobil |
| Fusion rocket | Research, development | Fast interplanetary travel, with limited Interstellar applications |  |
| Ground-level power supply | Standardisation, commercialisation | Reduction of required battery size and weight for battery electric vehicles | Trafikverkets Program för Elvägar |
| Hoverbike | Working prototypes, early commercialisation | Package delivery, search and rescue |  |
| Hovertrain, Ground effect train | Research, development | Trains with higher speed | Aérotrain, Duke Hospital PRT, Hovercraft |
| High Altitude Platforms | Experimentation | Communications |  |
| Jet pack or backpack helicopter | Early commercialisation, prototypes | More effective transportation |  |
| Maglev train, Vactrain | Research, early commercialisation | Trains with higher speed | Transrapid, Shanghai Maglev Train, Linimo |
| Magnetic levitation | Research, development, Commercialisation (Maglev Train) | High-temperature superconductivity, cryogenics, low-temperature refrigerators, superconducting magnet design and construction, fibre-reinforced plastics for vehicles and structural concretes, communication and high power solid-state controls, vehicle design (aerodynamics and noise mitigation), precision manufacturing, construction and fabrication of concrete structures, maglev car, maglev based spacecraft launch. | Vactrain, Levitra |
| Mass driver | Prototypes |  |  |
| Personal rapid transit | Early commercialisation, diffusion | More effective transportation | Morgantown PRT, ULTra |
| Photonic laser thruster | Prototypes |  |  |
| Physical Internet | Research |  |  |
| Scooter-sharing system | Commercialisation | Increased density | Bird (company) |
| Vactrain | Research, development | A faster way to get somewhere | ET3 Global Alliance, Hyperloop |
| Propellant depot | Research, development | enabling deep-space missions with more massive payloads and satellite life extension, ultimately lowering the cost per kg launched to space |  |
| Pulse detonation engine | Testbed demos | Fast interplanetary travel, with some possible interstellar travel applications |  |
| Self-driving car | Research, development, early commercialisation | Reducing concerns of tiredness while driving and also looking outside in the car. Helpful in countries where the employment of personal drivers is expensive. | Waymo, Tesla FSD |
| Space elevator | Research, development |  | Non-rocket space launch, Orbital ring, Skyhook, Space fountain |
| Spaceplane | Research, development | Hypersonic transport | A2, Skylon |
| Vehicular communication systems | Research and development, some diffusion | Vehicle safety obstacles inform others of warnings on entering intersections, traffic management, accommodating ambulances, fire trucks, and police cars to specific situations such as hot pursuits and lousy weather, driver assistance systems, and automated highways. | Artificial Passenger, Dedicated short-range communications, Intelligent transportation system |

**“Influence of IoT on the transportation industry”**

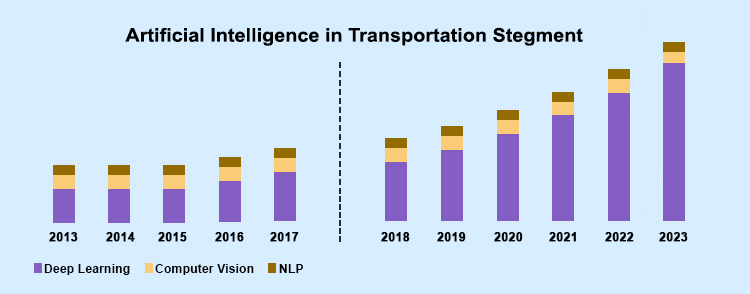
**Figure 1.2.**

**“Light Detection and Ranging”**

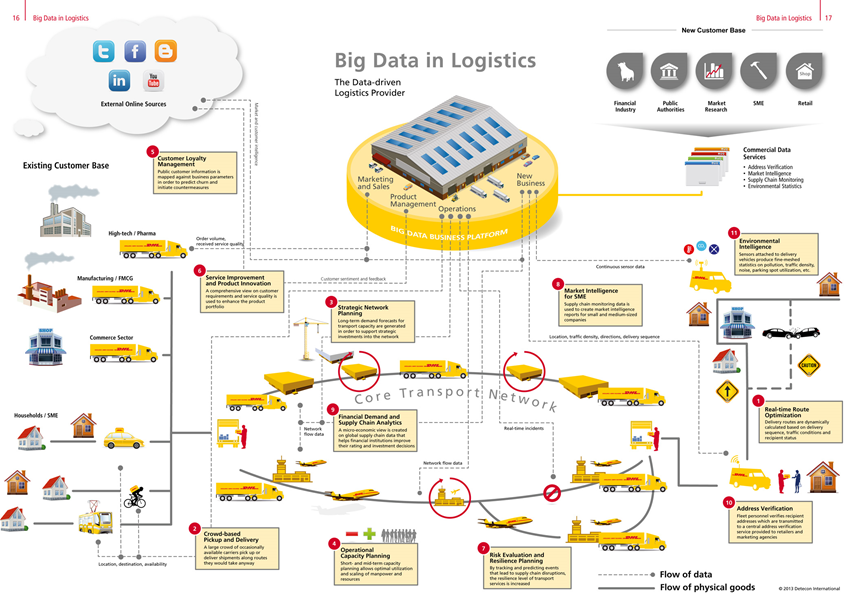


**Figure 1.3.**

**“Artificial Intelligence in Transportation Segment”**



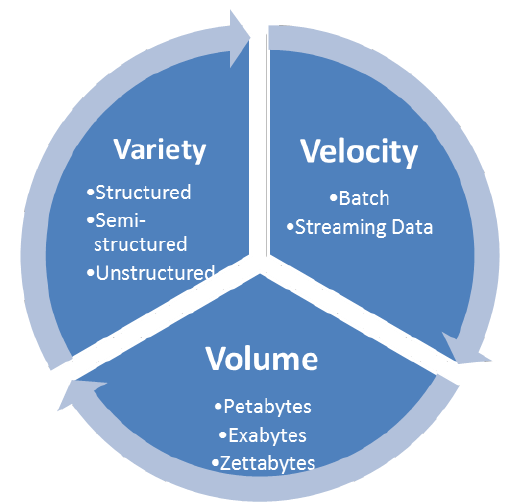
**Figure 1.4.**

**“Big Data in Logistics”**

**Figure 1.5.**

**“Big data in transportation business”**

**Figure 1.6.**

**“3Vs of Big Data”**

**Figure 2.1.**

**Table 3.1.**

**Big Data approaches used in ITS application areas**

|  |  |
| --- | --- |
| **Big Data approach** | **Problem/application (year)** |
| *Driver assistance and instrumented vehicles* | |
| multi-sensor data fusion for instrumented vehicles (2012) | Big Data fusion |
| efficient vehicle design (2015) | Big Data Analytics |
| driving data fusion techniques (2016) | Big Data fusion |
| RDMP framework for ADAS (2016) | Big Data platform |
| driving tendency recognition method (2016) | Big Data Analytics |
| behaviour and vehicle dynamics risk analysis (2016) | Big Data framework and policies |
| mobile agents for data management in vehicular networks (2017) | vehicular networks |
| Big Data schemes in social transportation systems (2016) | Big Data social transportation |
| guidelines to pioneer public transport (2016) | Big Data services |
| roadway control environmental footprint (2016) | Big Data Analytics |
| traffic congestion on limited access roadways (2016) | Big Data Analytics |
| road traffic operation (2016) | Big Data |
| traffic flow prediction based on deep learning (2015) | Big Data predictive analysis |
| R.C. evolution patterns (2015) | Big Data real-time analysis |
| O.D. matrix generation (2017) | Big Data Analytics |
| route planning services optimisation (2016) | Big Data Analytics |
| bus planning (2016) | Big Data Analytics |
| general traffic planning using IoT (2016) | Big Data Analytics |