



## Article

# Adaptive Cloud-Based Big Data Analytics Model for Sustainable Supply Chain Management

Nenad Stefanovic <sup>1,\*</sup>, Milos Radenkovic <sup>2</sup>, Zorica Bogdanovic <sup>3</sup>, Jelena Plasic <sup>1</sup> and Andrijana Gaborovic <sup>1</sup>

<sup>1</sup> Faculty of Technical Sciences Cacak, University of Kragujevac, 32000 Cacak, Serbia; jelena.plasic@ftn.edu.rs (J.P.); andrijana.gaborovic@ftn.edu.rs (A.G.)

<sup>2</sup> School of Computing, Union University, 11000 Belgrade, Serbia; mradenkovic@raf.rs

<sup>3</sup> Faculty of Organizational Sciences, University of Belgrade, 11010 Belgrade, Serbia; zorica@elab.rs

\* Correspondence: nenad@ftn.edu.rs

**Abstract:** Due to uncertain business climate, fierce competition, environmental challenges, regulatory requirements, and the need for responsible business operations, organizations are forced to implement sustainable supply chains. This necessitates the use of proper data analytics methods and tools to monitor economic, environmental, and social performance, as well as to manage and optimize supply chain operations. This paper discusses issues, challenges, and the state of the art approaches in supply chain analytics and gives a systematic literature review of big data developments associated with supply chain management (SCM). Even though big data technologies promise many benefits and advantages, the prospective applications of big data technologies in sustainable SCM are still not achieved to a full extent. This necessitates work on several segments like research, the design of new models, architectures, services, and tools for big data analytics. The goal of the paper is to introduce a methodology covering the whole Business Intelligence (BI) lifecycle and a unified model for advanced supply chain big data analytics (BDA). The model is multi-layered, cloud-based, and adaptive in terms of specific big data scenarios. It comprises business process modeling, data ingestion, storage, processing, machine learning, and end-user intelligence and visualization. It enables the creation of next-generation BDA systems that improve supply chain performance and enable sustainable SCM. The proposed supply chain BDA methodology and the model have been successfully applied in practice for the purpose of supplier quality management. The solution based on the real-world dataset and the illustrative supply chain case are presented and discussed. The results demonstrate the effectiveness and applicability of the big data model for intelligent and insight-driven decision making and sustainable supply chain management.



Academic Editor: Reza Eslamipoor

Received: 18 November 2024

Revised: 30 December 2024

Accepted: 1 January 2025

Published: 6 January 2025

**Citation:** Stefanovic, N.; Radenkovic, M.; Bogdanovic, Z.; Plasic, J.; Gaborovic, A. Adaptive Cloud-Based Big Data Analytics Model for Sustainable Supply Chain Management. *Sustainability* **2025**, *17*, 354. <https://doi.org/10.3390/su17010354>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** big data; sustainable supply chain; business intelligence; analytics; machine learning; model; cloud computing

## 1. Introduction

Supply chain networks are becoming increasingly complex due to the fast-changing and global business environment. Organizations are forced to embrace new business models, information sharing, integration and coordination, sustainability initiatives, and these new business practices turned out to be crucial for the global supply chain performance.

Besides traditional SCM practices, supply chains need to consider various social and environmental factors and also efficiently manage quality, risks, and waste. Supply chain sustainability can only be achieved by coordinated efforts of all participants and throughout the entire lifecycles of products and services. Collaboration between supply

chain participants on sustainability issues can reduce costs, risks, and waste, foster innovations, and improve branding and customer satisfaction. UN Global Compact [1] supply chain sustainability guide provides concrete steps and actions organizations should take to achieve supply chain sustainability. It introduces the continuous lifecycle model for supply chain sustainability that consists of the following phases: commit, assess, define, implement, measure, and communicate. The model incorporates principles of transparency, engagement, and governance to each phase and gives special emphasis on supply chain risks, productivity, and growth. Transparency and information-based decision making can contribute to less waste, optimized processes, and ultimately greener supply chains [2].

Sustainable initiatives and practices need to make their way into the entire supply chain. More than 70% of sustainability comes from suppliers, so managing those links and supplier-related processes can significantly improve supply chain performance. Some of the main actions supply chains can take include defining global and organizational sustainability goals, coordinated management, and information integration and exchange [3].

Sustainable supply chains imply data integration from various sources and data analysis, as well as digital collaboration platforms, to achieve a complete view of supply chain processes and to establish a more effective decision-making framework [4]. Supply chain intelligence is considered to be one of the key enablers and drivers for sustainable supply chains. This includes data integration, data processing into valuable information and knowledge, advanced analytics, and information delivery to decision-makers to take timely and optimal actions.

SCM enterprise systems are undergoing extensive transformations as organizations further depend on collaborative business communities to execute complex supply chain processes. Being such complex networks, supply chains typically generate enormous volumes of data that are demanding to integrate and analyze. While supply chains are becoming more and more complex, transiting from linear configurations to interrelated multi-level cooperative systems of organizations, there are ever-increasing amounts of data that need to be extracted, stored, managed, and analyzed. Business intelligence technologies and solutions have proven to be one of the best ways to successfully analyze such compound systems.

Supply chain intelligence is the application of BI methods, technologies, and best practices to better understand end-to-end supply chain operations by carrying out different types of analysis on data from organizations, supply chain participants, and other internal/external data sources (websites, government, social networks, devices, sensors, etc.). This comprises collecting, processing, analyzing, and managing large volumes of diverse data about customer relationship management, service, manufacturing, quality management, order fulfillment, supplier management, transportation, returns, e-commerce, etc. The goal is to facilitate effective operational, tactical, and strategic business decision making and provide the organizations with potential actions to improve supply chain performance [5].

Over the past several years, the way in which organizations integrate, analyze, deliver, and share their data has been altered significantly. Supply networks need to be more adaptive, responsive, and agile, which implies the effective management of rapidly growing volumes of various data and more intelligent decision making [6].

The current global and competitive business environment calls for timely and optimal supply chain decisions with a shortened analytical cycle—from data capture to insights and actions. This can be a very challenging task since supply chains typically produce vast amounts of data, including enterprise resource planning systems, geospatial data, large volumes of unstructured data (i.e., business documents, e-mails, blogs, and others), multi-

media (images, podcasting, and videos), social networks, and data created by numerous machines and sensor networks [7].

SCM business intelligence systems turned out to be incredibly effective in obtaining useful information and knowledge from various enterprise resource planning systems (ERP), but lately, many companies have to deal with new issues related to large volumes of data produced both within the supply chains and externally, data variety (various types of both structured and unstructured data), and data velocity (very often data are generated at high speeds and need to be ingested, processed and analyzed in the right time). Most present analytical systems are not capable of dealing with these new challenges [8].

The main challenges companies face when dealing with supply chain intelligence initiatives include the following [9]:

- Data silos and integration—Data in the supply chain is often scattered across various systems and companies, making it difficult to integrate and analyze comprehensively;
- Real-time data processing—Processing and analyzing data in real-time is crucial for timely decision making but can be technically challenging;
- Scalability issues—As supply chains grow and become more complex, the systems used to manage them must scale accordingly;
- Technological integration—Integrating new technologies like big data, IoT, and blockchain with existing legacy systems can be complex and costly;
- Sustainability and environmental impact—There is increasing pressure to adopt sustainable practices and reduce the environmental impact of supply chains.

Conversely, there are significant advancements in technologies such as cloud computing, in-memory computing engines, NoSQL databases, query languages and frameworks, parallel data warehousing, Internet of Things (IoT), artificial intelligence, data mining, visualization, etc. The term big data encompasses many of these technologies, which provide efficient and cost-effective ways to ingest, store, and process enormous volumes of data to provide useful information and knowledge for supply chain companies.

Today, supply chains are accumulating vast amounts of disparate data. Most companies view data as a critical corporate asset and consider data analysis as a very important instrument for gaining a competitive advantage. Still, only some of them are able to successfully manage and process these data. Those companies who have successfully implemented big data and business intelligence in their supply chain management (i.e., Amazon, Walmart, Proctor and Gamble, DHL, Nestle, General Electric, etc.) were able to enhance efficiency, visibility, and decision making, as well as to optimize supply chain processes and improve performance better than competitors [10].

Business intelligence systems were seen as an ideal means for data integration and collaborative analysis. However, due to the proliferation of new technologies and market pressures, many traditional BI systems are being challenged by today's data volume, velocity, variety, veracity, and value (5Vs). Additionally, BI systems should be capable of handling big data, performing predictive analytics and collaborative decision making via any device, as well as providing self-service BI capabilities [11].

Supply chain organizations face several vital challenges related to data analytics [12,13], as follows:

- Data silos—Without a cloud-based model, data often remain siloed within different departments, hindering comprehensive analysis and decision making;
- Limited scalability—Traditional on-premises systems may struggle to scale up to handle large volumes of data, leading to inefficiencies;
- High costs—Maintaining and upgrading on-premises infrastructure can be costly and resource-intensive;

- Lack of real-time insights—Organizations may miss out on real-time analytics capabilities, which are crucial for proactive supply chain management;
- Security and compliance issues: Ensuring data security and compliance with regulations can be more challenging without the advanced security features offered by cloud providers.

In order to thrive in a highly concurrent and global business environment, supply chains need to be agile, responsive, and adaptable. This entails specialized cloud-based big data analytical solutions that allow companies to respond without delay and to proactively take optimal actions. These solutions should enable sustainable supply chain operations by utilizing pervasive, collaborative, and user-friendly BI systems.

Cloud-based platforms offer scalable resources that can handle large volumes of data from various sources in real time. This is crucial for managing the dynamic and complex nature of supply chains. Cloud-based models enable real-time data processing and analytics, which are essential for timely decision making and responding to supply chain disruptions. They facilitate better collaboration among supply chain partners by providing a centralized BI infrastructure and tools accessible to all parties. Finally, these models help in tracking and analyzing sustainability metrics, such as carbon footprint and resource usage, enabling organizations to implement more sustainable practices [14].

Considering the issues and challenges of contemporary supply chain analytics, the purpose of this paper is to

- Provide a critical analysis of existing and traditional BI methods, approaches, and platforms through systematic background research and literature review;
- Examine current trends by analyzing the latest trends and advancements in the application of business intelligence and big data analytics within supply chain management;
- Investigate the benefits and challenges associated with implementing BI and BDA in SCM, including improvements in efficiency, decision making, and overall supply chain performance;
- Explore how BI and BDA can support sustainable supply chain practices and contribute to sustainable goals;
- Introduce the methodology suited for big data analytics in SCM, as well as the accompanying supply chain analytics lifecycle model;
- Present the architecture of the supply chain cloud-based big data system and its main advantages and benefits;
- Demonstrate the applicability and effectiveness of the proposed BDA methodology, the model, and the analytical cloud platform.

This paper deals with supply chain business intelligence with an emphasis on big data analytics systems, technologies, and tools. It examines challenges and new advances in supply chain intelligence and gives relevant background research on big data developments related to SCM. The paper presents an iterative and incremental methodology and a unified model for SCM big data analytics, which encompasses the entire business intelligence lifecycle. The model incorporates a scalable, flexible, and multilayered architecture with various big data services, to achieve scalable, adaptable, and performant analytical systems. Finally, the illustrative big data case related to supplier quality management is presented to demonstrate the usefulness and advantages of the proposed model.

## 2. Literature Review

Over the last two decades, the new business climate characterized by globalization and ever-increasing competitiveness has forced organizations and business networks to heavily invest in SCM enterprise information systems and digital services, in order to stay competitive and achieve sustainability. Even though these systems proved to be critical for

supply chain performance, they did not deliver full business value. Those systems were mostly transactional with a focus on the operational business view. The main deficiency was related to the lack of end-to-end supply chain analytical features, which resulted in isolated decision making and nonoptimal actions.

Many existing approaches to BI proved to be inadequate when challenged with big data features: variety of data types, data integration, master data management (MDM), new data modeling and querying techniques, metadata management, costs, and adequate knowledge and skills [15]. The variety, velocity, and volume of big data have significantly influenced supply chain business intelligence and analytics. New trends and technologies such as real-time analytics, machine learning, and data science are now becoming an integral part of BI solutions [16].

Present supply chain BI systems face several challenges, as follows [17]:

- Data volume—supply chains generate huge volumes of data that come from numerous sources, and companies need specialized tools to store and process those data;
- Data variety—besides structured data, there are various forms and types of unstructured data;
- Data velocity—data are being produced and ingested at high speeds, which causes new challenges for storing and analyzing data;
- Right-time analysis—having the right information at the right time becomes a necessity and a form of competitive advantage. Still, most supply chains lack infrastructure, services, tools, and apps for real-time or right-time analysis;
- Management and operations—big data systems are one of the most complex IT systems that are very challenging to design, deploy, and operate. Supply chains require not only more simple, scalable, and flexible infrastructures and platforms but also less costly and easier-to-manage services, such as those available in the cloud.

Existing models and methodologies for supply chain management often fall short in addressing the specific needs of sustainability, as follows [18,19]:

- Lack of comprehensive sustainability metrics—Many traditional models focus primarily on economic efficiency and operational performance, neglecting environmental and social dimensions. There is a need for models that integrate economic, environmental, and social metrics comprehensively;
- Inflexibility and rigidity—Traditional supply chain models are often rigid and inflexible, making it difficult to adapt to the dynamic requirements of sustainability practices and regulatory changes;
- Data silos and integration challenges—Existing methodologies often struggle with integrating data from diverse sources, leading to fragmented insights. There is a significant gap in the integration of diverse data sources and the use of advanced analytics;
- Limited use of advanced analytics—Many traditional models do not leverage advanced analytics, such as machine learning, big data, and AI, which are crucial for predictive and prescriptive insights;

As supply chain management is becoming increasingly complex, organizations are employing big data analytics to manage and coordinate end-to-end supply chain processes. BDA can improve supply chain performance by supporting business model enhancements, facilitating improved operations, and improving tracking capabilities. With big data technologies, companies are now able to mine information from both internal and external, structured, and unstructured data sources and provide supply chain insights that were formerly unattainable.

Big data are not related to a single technology but rather a blend of existing and some new technologies that can provide valuable intelligence for more efficient and effective supply chain management. This requires the capabilities to collect, store, integrate, and analyze increasing volumes of diverse data, at the right speed, while providing collaborative right-time analysis and actionable insights for decision-makers. In order to successfully realize these capabilities, new infrastructure, services, models, apps, and tools are needed. This includes new breeds of data warehouses and other data stores, new ways to extract, transform, and combine data, and new ways to query, analyze, visualize, and deliver data and information [20].

Some supply chains have experienced positive effects of big data projects: improved customer service levels, more efficient order fulfillment, enhanced demand–supply balancing, better responsiveness, agility, and problem-solving, optimized supply chain operations, and improved integration and visibility throughout the supply chain. But most early adopters have been less successful in producing such positive results from their projects. Holistic, supply chain-wide big data initiatives and solutions can involve high risks and substantial costs, so they need to be carefully planned, coordinated, and executed.

Supply chain sustainability can be improved by using big data analytical systems to reduce risks and uncertainties [21]. By applying big data analysis with unstructured data, the supply chain can improve resilience and enable sustainability. One of the studies showed that big data analytics can enhance sustainable supply chain management of manufacturing supply chains [22]. In recent years, numerous frameworks have been developed to enhance sustainability management within supply chains. However, these frameworks often exhibit various limitations. To address these gaps, a comprehensive framework for sustainable supply chain management was introduced, encompassing six key dimensions: methodology, organization, stakeholders, maturity model, human resources, and technology [23].

Regardless of the recognized benefits of big data analytics, many supply chains have faced difficulties in adopting it. Studies show that the main obstacles include large investments, lack of suitable frameworks and methodologies, inadequate infrastructures and architectures, as well as issues with information delivery and decision making [24].

There are also examples of successful implementation of big data solutions that demonstrate process improvement opportunities in various supply chain areas, as follows [25]:

- Big data solutions support integrated supply chain planning by making more responsive networks through a better understanding of partners and customers;
- The Internet of Things can supply various real-time telemetry data that can expose process details, while machine learning can be used for making predictions and uncovering hidden trends and patterns;
- Big data solutions can also improve distribution by utilizing various data sources (GPS, weather, traffic, logistics, etc.) to dynamically plan and optimize delivery;
- Supply chain risks can be mitigated by adopting proactive planning;
- Big data and BI systems can be applied in various supply chain processes [26];
- Planning—The processes associated with balancing demand and supply, developing and communicating supply chain plans, performance management, and alignment with overall business strategy;
- Sourcing—This includes processes related to purchasing, inbound transportation, receiving, storage, and transfer of materials, semi-products, products, and services;
- Production—This involves processes related to engineering, production planning, shop-floor control, quality management, materials requirements planning, assembly, etc.;

- Delivering—The processes related to sales, order fulfillment, finished product warehouse management, shipping, etc.;
- Reverse logistics—This includes processes associated with returning defective products to suppliers, as well as receiving returns of finished products from the customers.

Zhan and Hua proposed an integrated analytical infrastructure to enhance supply chain performance by utilizing various big data sources and analytical services [27]. The results obtained from the sports equipment manufacturing company demonstrate how adequate big data analytical infrastructure can integrate information silos, provide an integrated viewpoint of the company's operations capabilities, and facilitate decision making by data visualization.

Surveys show that 64 percent of supply chain executives view big data technologies as game-changing, but also view them as the cornerstone for future supply chain management and optimization [28]. Furthermore, 97% of supply chain managers state that their supply chains would benefit from big data systems, but only 17 percent confirmed implementing big data analytical systems within supply chain operations [29]. The main reasons include a lack of concrete value propositions and mature methods, as well as high risks.

Supply chains that put specific big data systems into operation have attained many concrete benefits, as follows [30]:

- Improved customer relationship management;
- More agile and responsive supply chain;
- Enhanced supplier and customer relationships;
- Increased efficiency and performance of supply chain operations;
- Higher level of integration throughout the supply chain;
- Better production planning, execution, and quality management;
- Optimization of warehousing activities and inventory management;
- More effective decision making;
- Higher level of sustainability.

During the last several years, big data analytics has become one of the most discussed topics among academics and supply chain practitioners. By studying existing research developments, it is possible to assess current findings and establish future research and development directions. Based on the systematic literature review, Nguyen et al. developed a classification framework based on four research themes: SCM areas in which big data has been applied, the level of analytics, types of big data models, and techniques deployed [31]. The study shows that most research is focused on single companies without an integrated view of supply chain processes and without holistic utilization of modern technologies such as cloud computing, IoT, data science, and internet technologies. It provides an indication of a positive relationship between the adoption of cloud computing use in process/activity integration, technology/system integration, and the integration of supply chain participants.

Wamba and Akter give a comprehensive literature study of big data analytics in supply chain management [32]. They provide main research guidelines and identify areas where BDA could have most effects on supply chain operations. These include establishing new methods and models for the design and development of BDA systems, as well as concrete BDA applications for improved decision making. Waller and Fawcett stress the need for further research related to data science, predictive analytics, and big data, especially in areas such as inventory management, transportation, supplier and customer relationship management, and forecasting [33].

One of the most important steps is to identify the right supply chain processes for big data applications. Those processes are typically data-intensive and critical for supply chain operations. Big data systems need to add concrete value in terms of improved

performance based on better decision making. Big data analytics can be utilized to enhance demand/supply planning and management, sourcing build-to-stock and build-to-order products, production scheduling and execution, warehouse and inventory management, transportation, delivery, and return processes. The study conducted by Roßmann et al. implies that BDA can improve demand forecasts, reduce safety stocks, and improve supplier relationship management performance [34]. Niu et al. proposed an intricate supply chain demand forecasting method based on graph convolution networks that showed better accuracy compared to traditional demand forecasting algorithms [35]. Ali et al. used the balanced scorecard methodology and artificial intelligence approach with artificial neural networks to enhance the prediction accuracy of supply chain performance metrics [36].

Although many companies have leveraged big data analytics capabilities to improve supply chain performance, the means by which they have achieved these capabilities are not well documented in the existing literature. For this reason, a stronger focus needs to be placed on recognizing the capabilities required to extract information and knowledge from big data. Arunachalam et al. introduced a maturity model that is conceptualized around five dimensions of BDA capabilities, specifically data generation capability, data integration and management capability, advanced analytics capability, data visualization capability, and data-driven culture [37].

Big data technologies are also changing the way companies manage inventories throughout the supply chain. There are examples of innovative companies that are using data from various data sources combined with optimization and prescriptive systems for more automated and intelligent decision making regarding inventory management. These tools can provide right-time performance information and feedback on the effectiveness of certain strategies [38].

Besides many potential benefits, there are still several main challenges for effective data-driven supply chain big data analytics, as follows [39]:

- End users and customers have increasing expectations from big data analytics;
- BDA cost efficiency and optimization;
- Compliance, security, and risk management and monitoring;
- BDA systems need to provision supply chain traceability and sustainability;
- In today's unstable business environment, BDA systems should enable better supply chain agility, flexibility, and adaptability.

Another important aspect of successful big data solutions is collaborative capabilities. Connectivity and information sharing have proved to be some of the important factors for effective big data analytics and have ultimately improved supply chain performance [40]. For this reason, technologies such as web portals, enterprise social networks, mobility, and similar collaboration and communication services and tools are becoming critical for collaborative planning and decision making in supply networks.

BDA systems for supply chain management also require significant investment in cloud infrastructure and platforms, including servers, storage, computational resources, and analytical tools. Many organizations opt for cloud-based solutions to leverage scalability and flexibility. Cloud services offer pay-as-you-go models, which can be cost-effective compared to on-premises solutions. These services typically offer a higher level of availability and security. They provide the computational power needed to run complex data pipelines, data processing queries, and machine learning models. The upfront costs for setting up big data infrastructure can be substantial, including hardware, software licenses, and implementation services. Cloud-based solutions can help mitigate some of these costs by offering scalable pricing models. Hiring skilled data scientists and analysts, as well as training existing staff, adds to the overall cost but is crucial for maximizing the benefits of big data analytics [41].



On the other hand, BDA can significantly improve supply chain efficiency by optimizing inventory management, reducing lead times, and enhancing demand forecasting. By identifying inefficiencies and optimizing processes, supply chains can achieve substantial cost savings. Access to real-time data and advanced analytics enables better decision making, leading to improved supply chain performance and competitiveness. Additionally, improved supply chain visibility and responsiveness can enhance customer satisfaction by ensuring timely delivery and better service levels [42].

Even though there is a lot of hype around various data analysis approaches such as big data, data science, business intelligence, predictive analytics, and others, academic studies related to applications of these approaches in SCM are still emerging. More concrete and rigorous research in this area needs to be carried out. Most of the articles study the general capacities of big data in SCM, as well as potential applications and values. Regardless of the advancements in information technologies and evident demands from the industry, methodologies, models, and software solutions related to big data analytics are yet to be studied and examined.

All these aspects combined point toward a necessity for a holistic approach to big data supply chain analytics with a specifically defined model and platform with a concrete combination of business intelligence and big data technologies that provide necessary flexibility, scalability, integration, collaboration, and security along the supply chain.

In the following sections, the unified supply chain big data model, the software architecture for SCM big data systems, as well as the example of a big data analytical system built using the proposed approach are presented.

### 3. Methodology and Model for Big Data Analytics

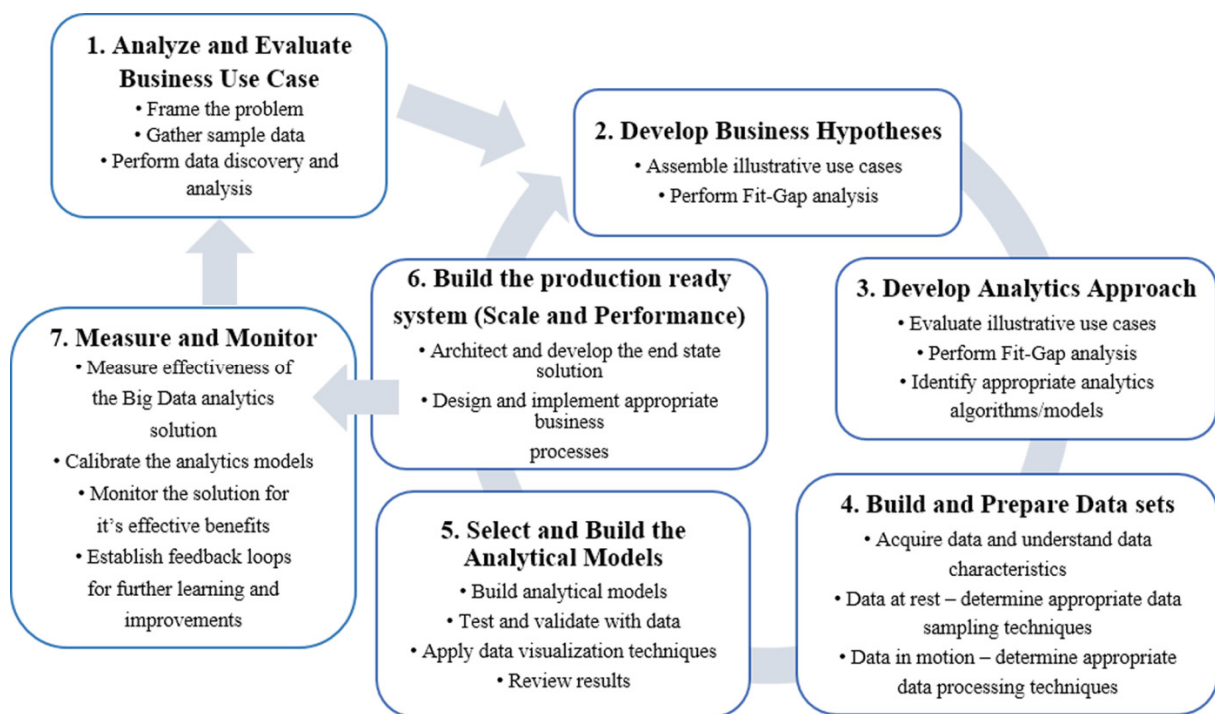
Although big data systems comprise many technologies and tools, the main emphasis of supply chain BDA is associated with questions, values, ideas, and innovations. They need to provide concrete and actionable insights to improve supply chain operations and sustainability. For this to be achieved, supply chains need to have a clear sustainability and business strategy and goals, adequate methodology for implementing BDA solutions, and the right combination of technologies.

#### 3.1. Supply Chain Big Data Methodology

Supply chain big data projects require careful planning and execution due to substantial complexity, high risks, regulatory requirements, and significant investments. They demand an approach that is different from traditional BI projects. When it comes to supply chain BDA, the challenges related to data heterogeneity and quality, timeliness, privacy, collaboration, and flexibility are even more expressed when compared to traditional BI systems.

Big data methodology usually involves a complex work breakdown structure with multiple phases with many subphases/subtasks that are carried out both sequentially and in parallel. The iterative and incremental method for designing and implementing BDA solutions, which comprises steps related to technology (infrastructure), data management (extraction, transformation, data modelling, etc.), and data delivery (reporting, visualizations, and collaboration) is proposed. Because of the complexity of supply chain processes, data speed, scale, and variety, an incremental method is proposed that involves progressive design, modification, and extension, instead of designing the entire BDA system at once [43]. The incremental nature of development cycles also plays a large part in reducing the risks related to BDA projects. This reduction in risk is perhaps one of the most significant factors for large companies that are skeptical about the potential business value of BDA projects.

Figure 1 shows an overall scheme of the proposed BDA methodology for executing iterative and incremental big data projects. After the use case analysis, steps 2 to 6 should be carried out iteratively through several design cycles. During each cycle, designers and analysts should incrementally include more data, apply various data modeling techniques, or add more analytical reports that are needed for solving particular supply chain problems. These iterations should ultimately result in an adequate solution that is agreed between designers, business analysts, stakeholders, and users. After deployment, these solutions should be periodically monitored for accuracy and quality and improved as needed.



**Figure 1.** Methodology for big data analytics.

The first phase is to analyze supply chain processes and evaluate the business case. This involves problem identification, specifications of business requirements, and initial data gathering and exploration. This should result in specific analytical goals and point toward the type of analytical solution that should be built (data exploration, predictive analytics, performance monitoring, stream analytics, etc.).

In the next phase, business analysts and designers should create illustrative use cases and carry out fit-gap analysis through several iterations, thus creating a detailed specification of the business hypothesis together with suitable infrastructure, data, and software needs.

The third phase includes defining appropriate analytical techniques/methods for solving supply chain problems. The type of analytical technique depends on the business problem, and it can be segmentation, prediction, association, what-if analysis, etc.

The next step is to create suitable quality datasets that will be used by analytical models and algorithms. This is a major challenge due to the complexity, heterogeneity, and scale of supply chain data. Data from the identified data sources (structured and unstructured) need to be extracted, transformed, and loaded into appropriate data stores (data warehouses, Spark-based datastores, events hubs, etc.). Metadata management is critical for big data projects, especially when transforming data. Supply chain metadata repository (centralized or distributed) ensures better data consistency and value extraction.

The following phase comprises building analytical models and developing prototype solutions using smaller data samples. This assumes experimenting and gradual refinement through several iterations. After the successful evaluation of the results, the production-ready system capable of handling the scale of big data should be established. Here, infrastructure and cluster configurations should be specified, along with a combination of big data tools and technologies for each analytical segment (ingestion, storing, analysis, and reporting). For supply chain BDA systems, it is critical to equip end users with intuitive and useful tools that enable them to explore data, gain insights, and take action.

The final stage is to monitor and measure the effectiveness of the BDA solution and make appropriate adjustments. Establishing the feedback mechanism is essential for continuous improvement in the BDA solutions.

The presented methodology differs from most existing BI methodologies because it encompasses the whole analytical life cycle, starting from project initiation and requirements analysis, all the way to system deployment, monitoring, and improvement. It features an iterative approach where knowledge from each iteration is used to improve system design in the next pass. The methodology is also flexible enough to support the design and development of various supply chain BDA systems.

In order to accommodate such flexibility and different types of processing tasks, it is necessary to define suitable big data models and supporting infrastructure that will enable the design and composition of various analytical applications.

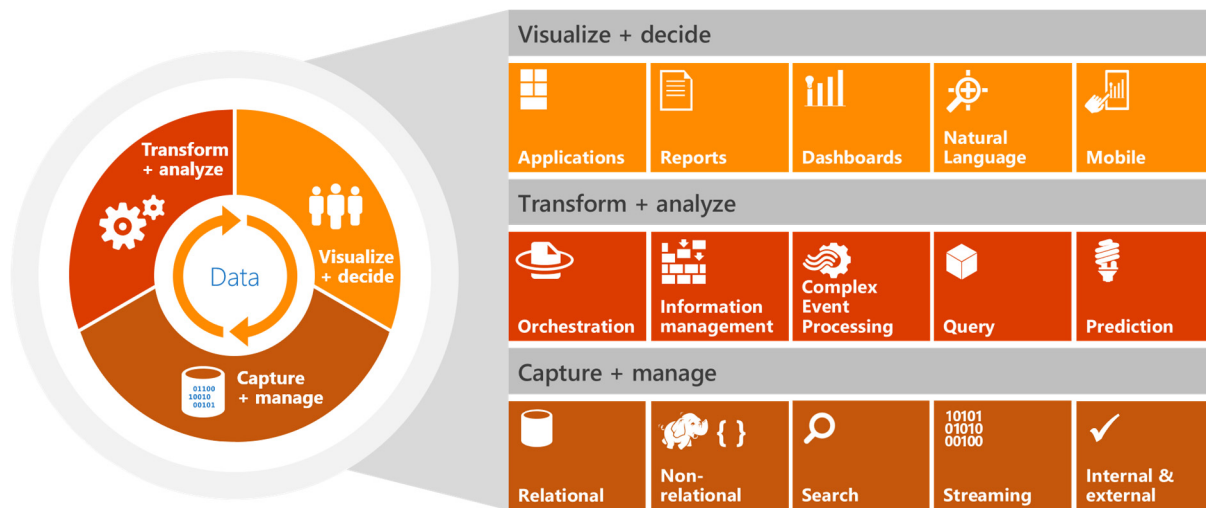
### 3.2. Supply Chain Big Data Lifecycle Model

Technologies cannot be undervalued since they are the key enablers for achieving the goals of supply chain big data systems, process optimization, and improvement, as well as innovations in products and services. Since supply chain processes are very complex and each organization can have specific requirements and technological constraints and needs, analytical requirements can vary significantly. Therefore, big data architecture should be layered (provide services for data ingestion, transformation, modeling, analytics, and delivery) and be flexible to support various analytical scenarios, as well as enable future additions.

To surpass the key shortcomings of present supply chain business intelligence systems and to solve the main challenges of big data analytics in SCM, a unified multi-layered supply chain BDA model is proposed. The model incorporates cloud-based infrastructure, platforms, services, and tools for data ingestion, ETL (extract, transform, and load), storage, processing, analysis, and reporting. It has been designed to integrate with existing business intelligence and collaboration systems. It is based on the supply chain process metamodel, with process metrics and best practices for actions and improvements. The model includes a flexible supply chain process, data modeling approaches, and a multi-layered system architecture that supports the design and development of composite BDA systems. Supply chain process modeling, metrics, and best practices are based on the standard supply chain operations model (SCOR), which provides necessary standardization and interoperability for analytical information systems. This way, supply chain participants can use the entire set of analytical platforms, which lets them use the suited services and tools for the right analytical job when developing big data systems.

Figure 2 illustrates the architecture along with the main layers and services.

As the model relies on a variety of cloud services to achieve its scalability and performance requirements, the choice of cloud provider is flexible (it is compatible with the analytical services of leading cloud providers); the model can also be deployed in a variety of environments (supply chain can decide to use public, private, or hybrid clouds).



**Figure 2.** Supply chain analytics lifecycle model.

One of the most important steps in the supply chain analytics lifecycle is data ingestion and preparation for further data processing and analysis. The data integration and analytics layer handles a variety of data types (relational, unstructured, streaming, OLAP (On-Line Analytical Processing), etc.) and data sources (ERP, SCM, B2B platforms, e-commerce websites, third-party logistics providers, sensors and IoT networks, etc.) via managed cloud ETL services. These services enable data-driven workflows and pipelines for data orchestration and transformation at scale. Data pipelines can be used to perform various data transformations and data flows through a range of predefined or custom-built connectors for various supply chain data sources and services and to design complex ETL workflows with advanced control flow logic that can move or refresh large volumes of data. Data pipelines can be also used for incorporating DevOps practices such as continuous integration and continuous delivery (CI/CD) workflows within the BI lifecycle model. This means that the process of development, testing, and deployment can be automated significantly, resulting in better data quality, faster delivery, and efficient decision making.

Data storage and processing are based on the massively scalable cloud Data Lake technology with supplementary services that enable more options for data querying and modeling. Data Lake includes HDInsight clusters as a managed service in the Kubernetes Service infrastructure. This enables various analytical engines (Spark, Storm, Kafka, HBase, etc.) to run side-by-side. Data can be queried by standard languages and data sciences frameworks, as well as with specialized U-SQL query language that enables the parallel execution of map-reduce style programs. Raw and existing supply chain data can be queried directly via DirectQuery technology, always providing current reports and eliminating the need for data import. The analytical services include various data analytical engines, models, and schemas. These involve data exploration, machine learning (predictions, segmentation, pattern recognition, etc.), and performance measurement models.

The visualization layer delivers insights to appropriate users via reports, dashboards, and advanced analytics such as self-service BI, natural language processing, performance management, and collaborative decision making. The main element of this layer is a dedicated business intelligence portal, which serves as an integrated analytical front-end, with various analytical and visualization services and features for planning, collaboration, sharing, and actions.

The introduced BDA model unifies methods, processes, services, and tools into the integrated big data system. Using the presented cloud services, supply chain member companies can gather, store, and manage data from various sources and in various for-

mats and then process and transform that data in accordance with specific analytical requirements. This enables integrated, timely, and transparent monitoring of operational and sustainability key performance indicators, as well as more optimal decision making and actions.

#### 4. Cloud-Based Big Data Solution—Results and Discussion

To demonstrate the applicability and effectiveness of the proposed supply chain big data model, the concrete cloud-based analytical system was designed. Furthermore, a supplier management solution based on the real-world dataset and using the described big data methodology was designed, developed, and deployed. The main results and advantages are further presented and discussed.

##### 4.1. Supply Chain Big Data Analytical System

The proposed supply chain BDA model requires the careful selection and composition of various analytical services in order to support a variety of analytical scenarios. Figure 3 shows the general architecture of the real-world big data system for supply chain analysis. The architecture is cloud-based (Azure platform) and provides necessary adaptivity, flexibility, extensibility, scalability, and security.

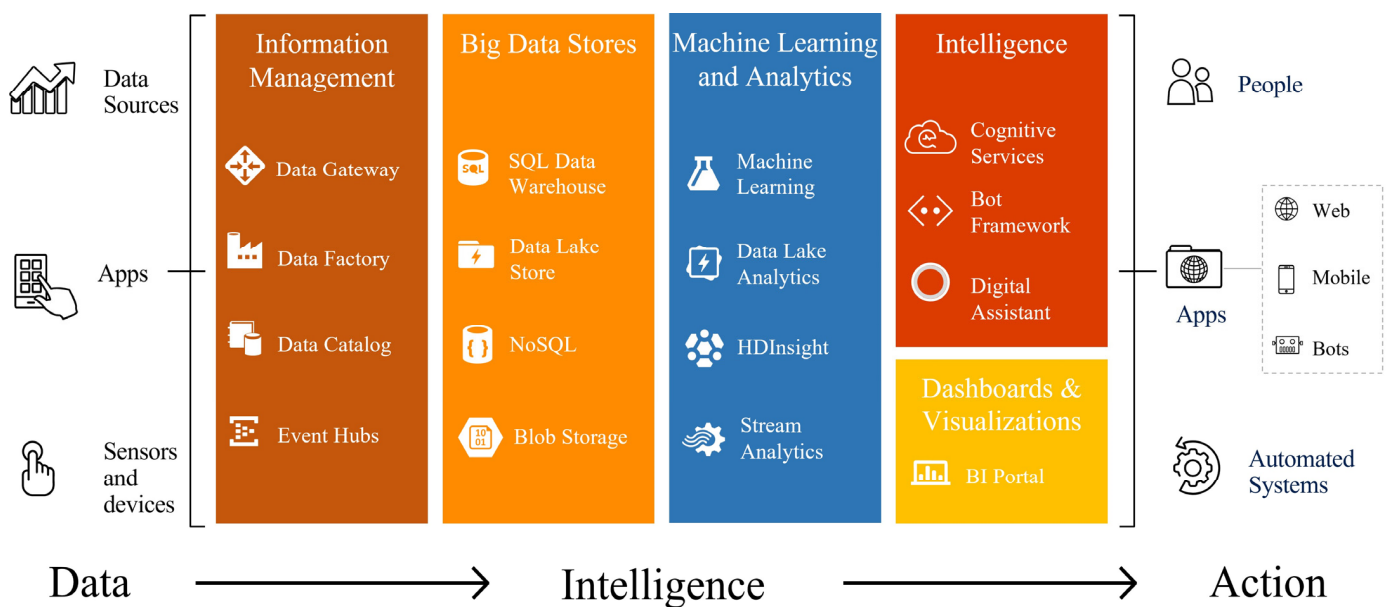


Figure 3. Supply chain big data solution architecture.

The first layer includes several cloud-based data management services such as data gateways, data catalogs, ETL workflows, and big data streaming. ETL jobs for data extraction, cleansing and import, and event hubs act as a scalable data streaming platform capable of ingesting large amounts of high-velocity data from sensors and IoT devices throughout the supply chain.

Event Hubs is a real-time data streaming platform-as-a-service with distributed architecture and low latency. It is based on the Apache Kafka engine and capable of ingesting, buffering, storing, and processing data streams in real time. This service utilizes a partitioned consumer model, which allows concurrent processing of many data streams. This is useful in supply chain scenarios where many companies can stream data simultaneously.

The Data Factory service enables the serverless creation of both ETL end ELT (Extract, Load, and Transform) data-driven pipeline workflows to ingest, prepare, transform, and process data from various supply chain data sources. Workflows and individual data

tasks can be saved as templates and reused as needed since supply chain participants can have matching or similar data integration scenarios. These reusable ETL workflows can be utilized across multiple semantic data models and reports by different supply chain organizations. Data pipelines can also be used for ingesting data into a single data lake, allowing it to be accessed by other cloud analytics services. This single source of truth can be utilized for integrated and coordinated analysis and decision making throughout the supply chain. Furthermore, through the Lakehouse approach, supply chain data sources are abstracted, which eliminates the need for direct access to data sources and facilitates access control and security management.

For scenarios where companies want to keep their data sources in-house, it is possible to install on-premises data gateways and connect them (register) to the single cloud data gateway via an encrypted connection. The cloud data gateway provides data to other services such as those for ETL (data pipelines) and analytics.

Furthermore, the global supply chain data catalog was deployed in the cloud. This is a specialized cloud-managed common data service that allows clients to discover, understand, consume, and extend data models. The data catalog is based on a crowdsourcing model of semantic metadata, annotations, entities, attributes, and relationships and enables all supply chain information workers to add their expertise to build suitable data models that can be reused and integrated into specific analytical applications and services. The idea is to achieve structural and semantic consistency and to simplify the integration and disambiguation of data since supply chain data sources and schemas are typically diverse.

Data Factory pipeline workflows can further store data in various big data stores, depending on the specific data tasks and data formats: Data Warehouse, Data Lake, Blob storage, OneLake, or some other cloud-managed NoSQL storage. Data Lake is a highly scalable and performant service capable of storing data of any size, speed, and shape. It can store transformed or original data that can be later loaded into a data warehouse or some analytical service for processing. Blob storage can be used for unstructured supply chain data such as documents, images, audio, or video. It supports controlled data sharing, so each supply chain consumer can access only authorized data. OneLake is a unique, unified, and logical data lake ideal for supply chain data storage scenarios. It provides a single copy of data and data items (data lakes, data warehouses, etc.) that can be used by various analytical engines thanks to the open data format. OneLake also supports the shortcuts, which enable data fusion from various cloud systems and supply chain data sources, without the need for data copying and duplication.

Data Warehouse is a cloud-managed service that provides a single version of truth for the whole supply chain. Staged data from the Blob storage can be transformed and cleansed data using Databricks (cloud-based Apache Spark engine) and imported into the data warehouse. The main benefits include data integration (structured and unstructured), mature modeling methods and tools, and consolidated semantic data structure. This enables faster and easier data analysis and visualization.

Data from big data stores can be used directly for reporting and visualization, but it can also be used for advanced analytics. Data lake analytics enables the creation and execution of on-demand analytical jobs without special servers or virtual machines. It can consume data from the data warehouse, data lake, or Blob storage.

Machine learning services can be used to design, deploy, train, automate, manage, and track machine learning models. These models derive new knowledge via predictions, segmentation, pattern recognition, or associations and enable supply chain managers to make better decisions based on insights. Stream analytics is a complex event-processing engine whose purpose is to process large volumes of high-velocity streaming data from

multiple supply chain sources. The data can come from various supply chain telemetric systems, e-commerce websites, geospatial systems, social networks, etc.

Lastly, information and knowledge obtained from previously described analytical services and models should be supplied to information workers at the right time and in the right format. In this context, a specialized BI web portal is employed. The portal serves as a single point for data analysis, reporting, and collaborative decision making. For advanced analytics, BI portals and other analytical apps can be further enhanced with intelligent cognitive services such as natural language processing (NLP), computer vision, and speech processing or integrated with bot frameworks for automatic or semi-automatic planning and decision making in the form of digital assistants.

#### *4.2. Illustrative Supply Chain Supplier Management Big Data Analytical Solutions*

Sourcing costs typically account for more than 70% of total costs in industry-related supply chains. Analytical systems can be highly beneficial for more efficient and effective supplier management. Tracking supplier performance and product/parts quality can lead to process optimization, higher quality throughout the supply chain, and significant cost reduction.

To demonstrate the proposed big data methodology, model, and system architecture, a supply chain big data analytical solution supplier relationship management analysis was designed. An anonymized supplier relationship management dataset from a real-world supply chain was used [44]. It contains data from 24 manufacturing factories and from several states. Additional data were integrated from various available sources (i.e., maps).

The proposed iterative and incremental big data methodology was used to manage all the steps required for such an analytical solution, as follows:

- Analysis and evaluation of supplier relationship management use cases;
- Developing the business hypothesis with illustrative use cases and data exploration and gathering;
- Identifying appropriate analytical techniques, methods, and tools for solving supply management analytical requirements;
- Design of the big data architecture and implementation in the cloud environment;
- Data preparation, extracting, cleansing, transformation, and loading;
- Building the concrete analytical models, testing, and validation;
- Information visualization through the design of the dashboards and reports;
- Deployment of models and reports to the cloud analytical services, including the data warehouse and the BI portal;
- Monitoring and evaluating the effectiveness of the solution and providing feedback for further improvements.

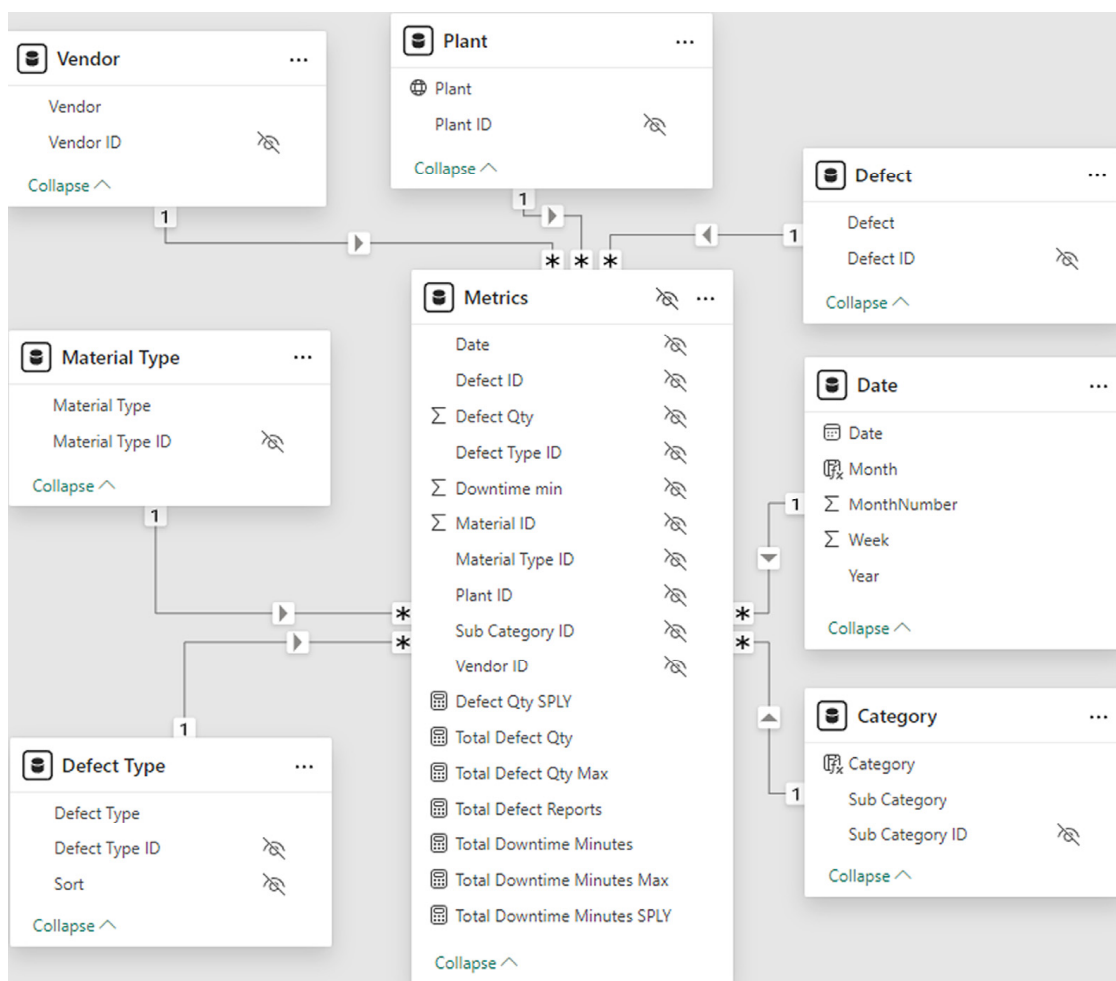
Based on the presented supply chain analytics lifecycle model and the general big data solution architecture, the concrete big data solution that includes data ingestion, transformation, storage, analytical processing, and visualization was designed.

Data from several data sources such as relational databases, files, and web APIs, is ingested via cloud Data Factory workflows into the Data Lake Store in the form of related entities of the Data Catalog (Common Data Model). The main dataset with operational data is retrieved via on-premises and cloud gateways, with auto-refresh capability. Using the Data Factory jobs, data from the data lake store is then transformed, cleansed, and loaded into the cloud-managed Azure Data Warehouse with the predefined multidimensional data model, including dimensions, measures, cubes, and key performance indicators (KPIs). The data warehouse serves as a centralized and integrated supply chain analytical data store. This ensures the required data quality, reporting consistency, and better performance.

Using the Azure Databricks (Apache Spark-based analytics engine) workflows, data from the Data Warehouse are extracted, prepared, and sent to cloud machine learning services for processing and then to the cloud-based business intelligence service for advanced insights.

BI service and the BI web portal are used for information visualization from the Data Warehouse and Databricks and collaborative analysis and decision making. They support self-service BI and integration with cloud cognitive APIs (Application Programming Interfaces) and services, such as those for automated decision making and language processing.

Considering the specifics of the supply chain (different companies, data sources, inability to alter existing systems, etc.), an abstracted multidimensional data model with dimensions, metrics, and key performance indicators was created. This data model is ideal for supply chain analytical scenarios, and it is part of the unified BI semantic model [11]. Figure 4 presents the segment of the multidimensional data model, one of the analysis services cubes, with dimensions, facts (metrics), calculations, and KPIs.



**Figure 4.** Supplier quality analysis services cube.

This kind of data structure enables better performance due to a star schema, pre-processed aggregations, and an in-memory column store engine. Furthermore, the rich semantic model ensures modeling consistency for the supply chain and provides a unified base for end-user reporting where different systems (BI portals, desktops, or mobile apps) can consume analytical artifacts from a single place. It is also suitable for other analytical tasks, such as self-service BI or machine learning experiments.



Different cubes (models) can be designed to fit specific analysis tasks and to facilitate access control and authorization. The same dimension/measure/KPI can be used in different cubes, thus providing master data management throughout the supply chain. Fact entity contains various metrics (related to downtime and defects) and dimensions (date, plant, vendor, category, etc.). The data model enables various analyses and performance monitoring related to supplier quality and plant operations (defects and downtime). This way, many of the existing BI solutions or artifacts can be reused or directly migrated to the cloud for better performance and manageability.

A variety of analytical reports are designed using various business intelligence technologies, services, and tools. The most important BI artifacts (dashlets, i.e., reports, spreadsheets, charts, maps, etc.) are combined within several dashboard web pages. Figure 5 shows the dashboard page with various dashlets, which present information in various forms and from various data sources.

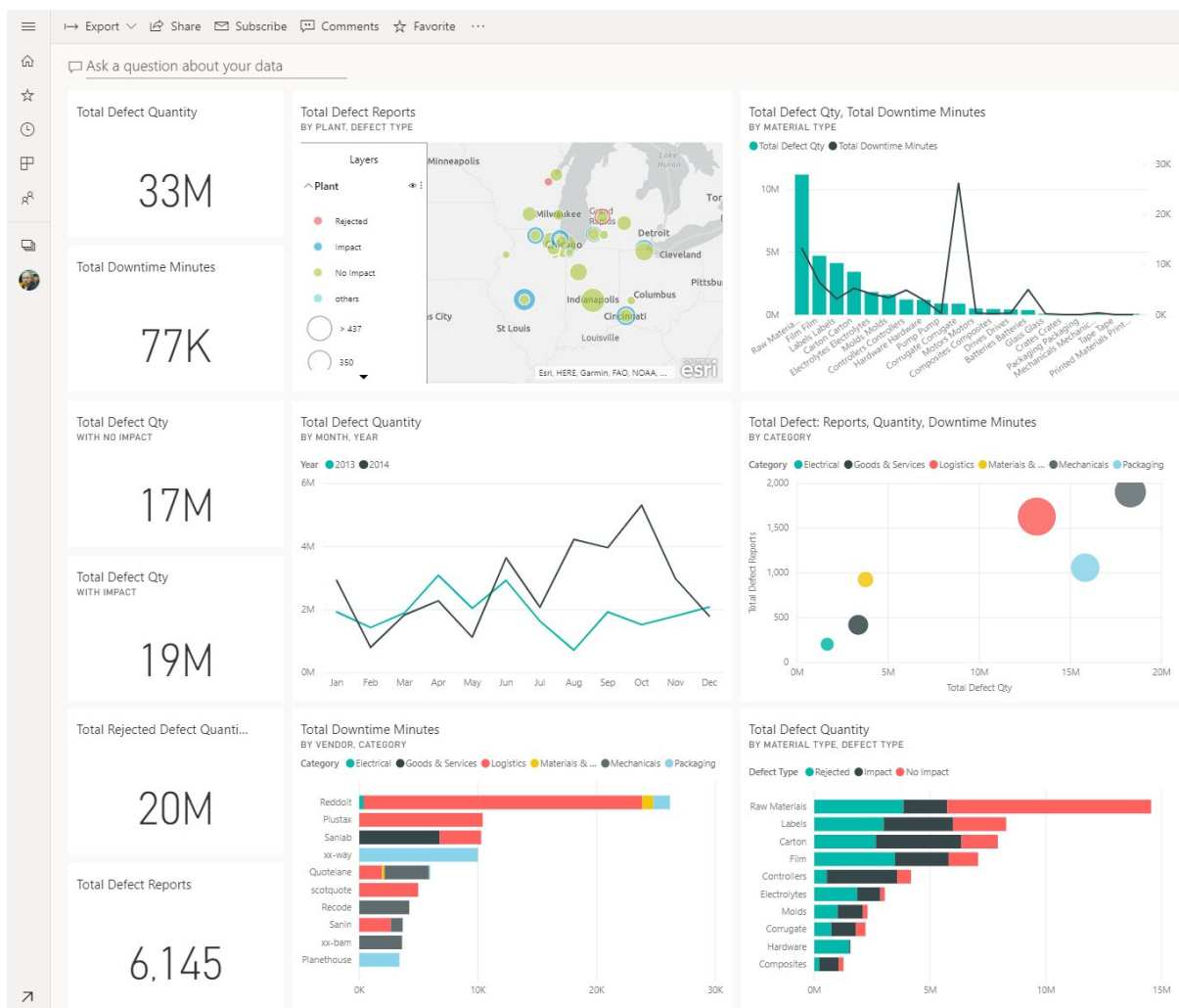


Figure 5. Supplier quality dashboard page.

As a result of automatic data refresh and high-speed querying (high-speed in-memory storage), it is possible to monitor actual KPIs and take proactive actions. Reports and dashboards can be adequately displayed on mobile devices, thus enabling users to track information and from any device and any place.

The dashboard is designed with dashlets and scorecards that are associated with drill-down reports to provide further information exploration and analysis and ultimately better information-based decision making. For instance, if an employee needs to analyze

how factories manage and handle faulty materials and interruptions, by selecting the map dashlet (as well as some other web page segments), it opens another dashboard page (shown in Figure 6), which can be applied for further drill-down analysis and filtering in order to obtain valuable knowledge about supplier quality and production management and take corrective actions based on the SCOR best practices.

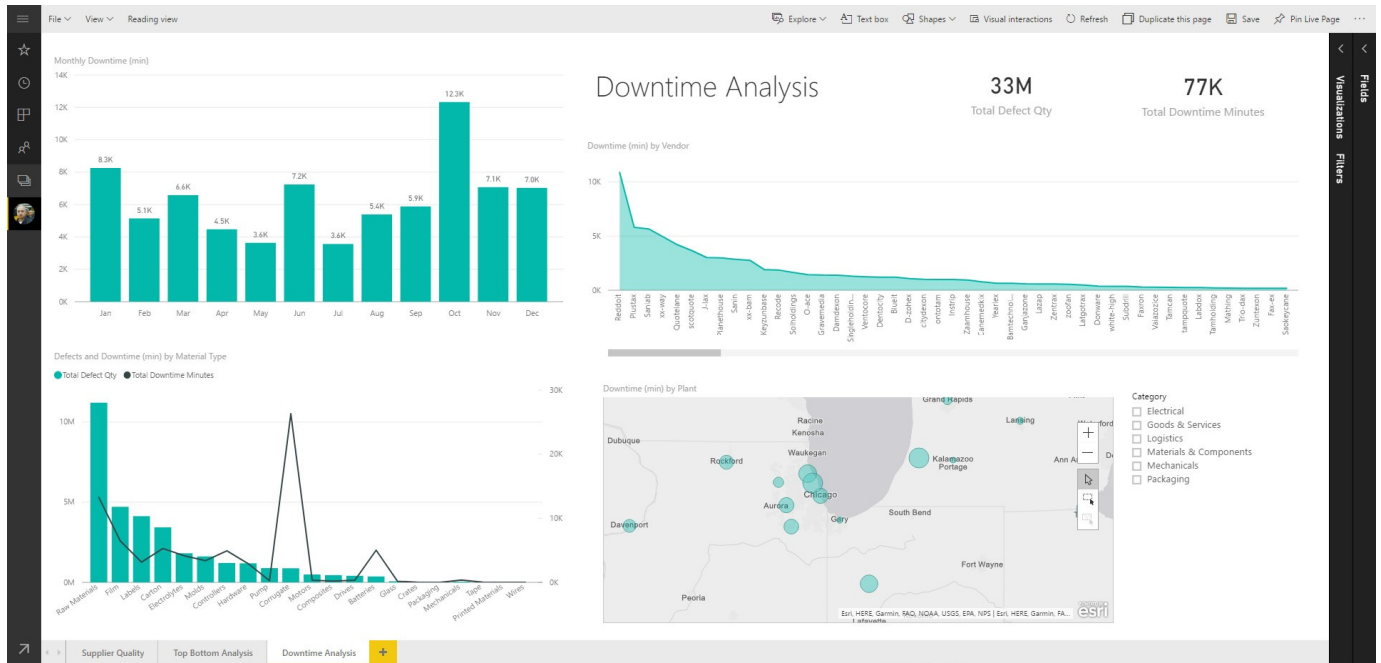


Figure 6. Supplier downtime analysis dashboard.

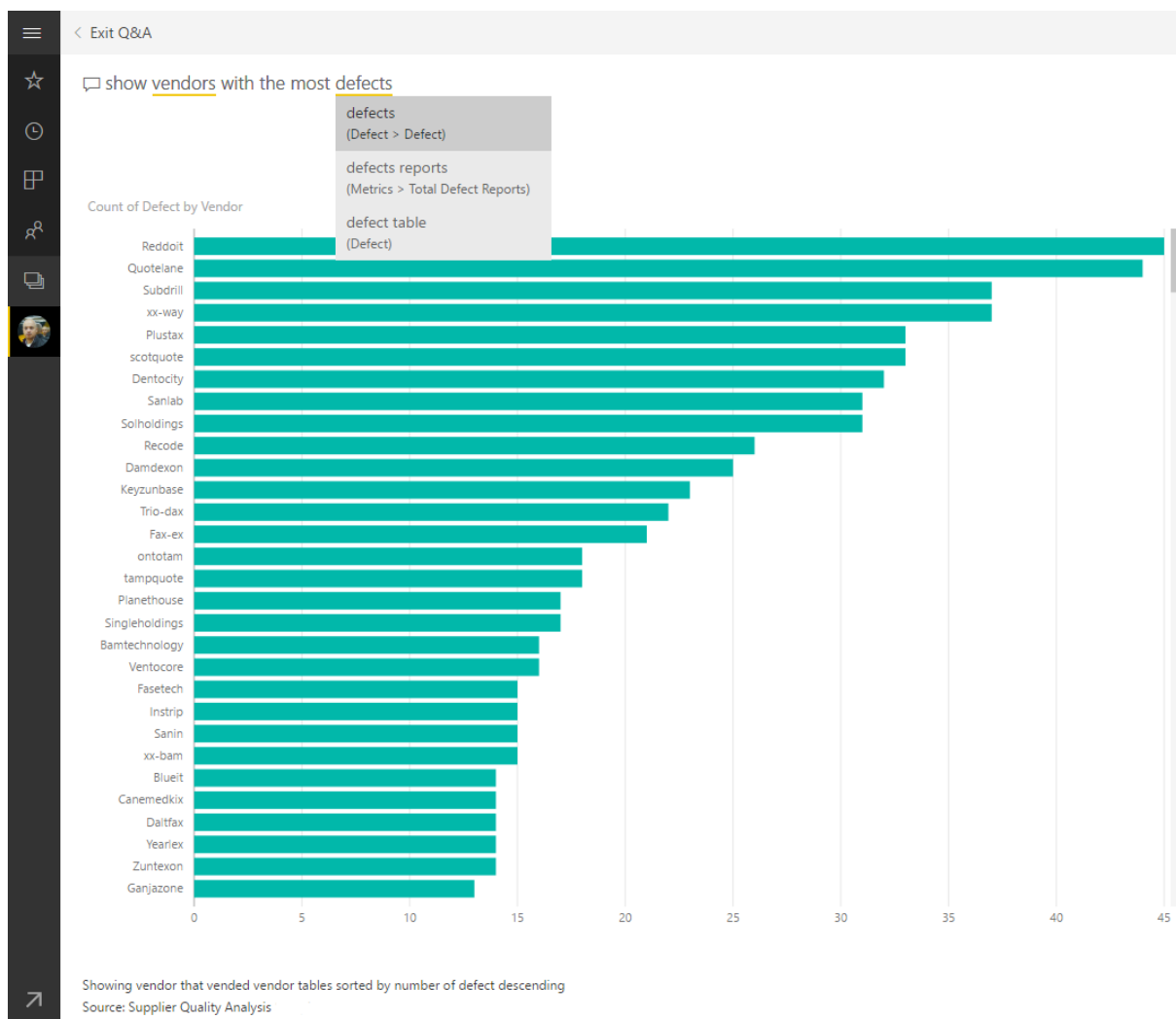
The report consists of various types of dashlets, and it is designed in such a way that quality managers can quickly obtain general insights. For instance, the analysis shows that some materials have fewer defective parts but they can cause a huge delay resulting in larger downtime. By analyzing the total downtime minutes information and the downtime by type of material, it can be observed that corrugated materials have fewer defects but cause the most downtime, whereas raw materials have the most defects and also cause significant downtime. Additionally, by selecting the Corrugate column in the chart, users can see which factories are impacted most by this defect and which vendors are most responsible. Supplementary analysis can be performed by selecting a specific factory in the map dashlet to understand which vendor or material is responsible for the interpretations at that factory.

Additionally, the monthly downtime analytical chart shows monthly downtime where information can also be dynamically filtered by selecting a specific month (by filtering a single chart, other chart data on other charts is automatically filtered). From the chart, it can also be seen that the most downtime has been in October and that the raw materials cause the most downtime. This information can be associated with the supplier that delivered most of the raw materials that month. Quality managers can inform the supplier, which can further analyze its processes and thus collaboratively analyze problems and perform actions. Furthermore, information can be observed by category parts, year, and defect type.

Most of the existing BI systems do not provide users with intuitive, flexible, and customizable analytics. The presented big data system provides advanced self-service BI capabilities. Thanks to the designed multidimensional data model, it is possible to dynamically filter data by various dimensions (i.e., defect type, plant, vendor, year, process

category, etc.) and change existing dashlets (select different dimension attributes, metrics, and KPIs or change visualizations).

Also, this data model enables users to ask questions using natural language queries as shown in Figure 7. The system uses a data schema and the cloud-based natural language processing API to intelligently filter, sort, aggregate, group, and display data based on the keywords in question (underlined in yellow color), thus enabling more user-friendly and intelligent interaction. The results can be saved and pinned to the dashboard, so knowledge can be shared with other employees and partners.



**Figure 7.** Result of natural queries.

Based on the analysis, users can share certain reports or dashboards with co-workers in charge of a particular plant, department, or process. The presented BI portal is integrated with the enterprise collaboration platform, so users can easily find relevant persons, share information, communicate (chat, video), collaboratively analyze processes, and take necessary actions.

Additionally, users can subscribe to relevant reports or dashboards and be notified by e-mail at certain times. It is also possible to create automatic alerts so that users are immediately notified when certain metric passes the defined threshold.

The data model and the BI portal framework enable further development and extensions to the supplier quality analysis. For example, it is possible to add new reports or

create even machine learning models (upon existing multidimensional models) for making predictions, pattern recognition, etc.

Even though the presented big data analytical solution provides advanced reporting, self-service BI, and exploratory analysis, additional knowledge can be derived by incorporating machine learning models. For this purpose, several machine learning models are designed to provide more advanced insights, which include discovering hidden patterns, trends, outliers, and key influencers, as well as correlations and predictions. Machine learning models can be applied for the entire dataset, or scoped for the particular dashlet (tile), so the user can gain more focused insights. Figure 8 shows segments of the automatically derived knowledge from the supplier dataset.

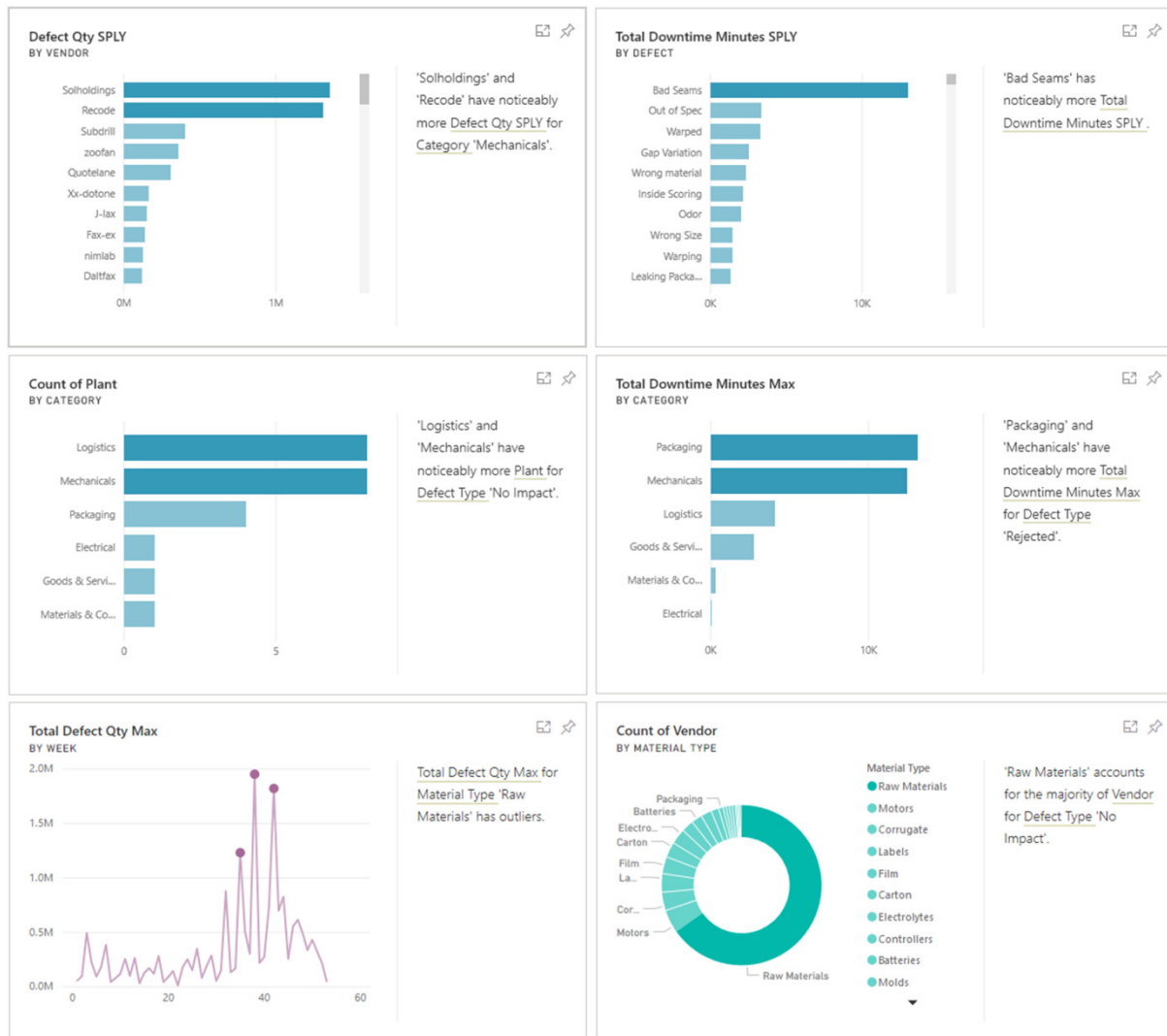


Figure 8. Automatically generated machine learning insights.

Based on the machine learning models, different types of new knowledge can be obtained. For example, it can be noticed that two vendors have noticeably more defects in the Mechanical category. The Logistics and Mechanical categories have noticeably more factories when it comes to the “No Impact” defect type. The “Bad Seams” type of defect caused most of the downtime. The maximum downtime was for defect type “Rejected” and for “Packaging” and “Mechanical” categories. The time series outliers machine learning model detected the total defect quantity outliers for the raw materials. Raw materials accounted for the majority of vendors for the “No Impact” defect type.

Decision-makers can use this derived knowledge and insights to make smarter decisions and take appropriate actions.

In order to make the design more efficient and to allow reuse, specialized BI packs have been created. These BI packs are special types of apps that contain datasets, workbooks, reports, and dashboards. These elements can be published to the supply chain-wide BI pack library and securely shared and reused throughout the supply chain, as well as customized or extended. This enables faster deployment, consolidated reporting, and single-point maintenance.

It is possible to create different workspaces (with reports, dashlets, dashboards, datasets, spreadsheets, etc.) for specific analytical processes and business segments (department, company, or supply chain), where developers and business users can collaboratively design BI artifacts. These artifacts can also be embedded (reused) via web portal API by supply chain companies in their own information systems, thus creating flexible composite BI apps.

The presented BDA supplier management solution can be used for more informed, real-time, knowledge-based, and predictive decision making. Managers can take timely and proactive actions to decrease costs, minimize defects and production downtime, and monitor suppliers' performance.

Implementing the proposed cloud-based big data analytics model can significantly transform supply chain management, making it more efficient, sustainable, and resilient. The main practical implications for sustainable supply chain management include the following:

- Enhanced decision making
  - Implication: Real-time data processing and advanced analytics enable more informed and timely decisions.
  - Benefit: Organizations can quickly respond to supply chain disruptions, optimize inventory levels, and improve overall efficiency.
- Cost efficiency
  - Implication: Cloud infrastructure reduces the need for significant upfront investments in hardware and maintenance.
  - Benefit: Lower operational costs and the ability to scale resources as needed without large capital expenditures.
- Improved collaboration
  - Implication: A centralized cloud platform facilitates better data sharing and collaboration among supply chain partners.
  - Benefit: Enhanced coordination and communication lead to more synchronized and efficient supply chain operations.
- Sustainability tracking
  - Implication: The model provides tools to monitor and analyze sustainability metrics.
  - Benefit: Organizations can implement more sustainable practices, reduce environmental impact, and comply with regulatory requirements.
- Scalability and flexibility
  - Implication: Cloud-based solutions offer scalable resources that can handle large volumes of data from various sources.
  - Benefit: The ability to adapt to changing business needs and scale operations efficiently as the supply chain grows.
- Risk management

- o Implication: Advanced analytics can predict potential risks and disruptions in the supply chain.
- o Benefit: Proactive risk management strategies can be developed, reducing the impact of unforeseen events.
- Enhanced customer satisfaction
  - o Implication: Improved supply chain visibility and efficiency lead to better service levels and faster delivery times.
  - o Benefit: Higher customer satisfaction and loyalty due to reliable and timely product availability.
- Data-driven innovation
  - o Implication: Access to comprehensive data analytics fosters innovation in supply chain processes and strategies.
  - o Benefit: Continuous improvement and the ability to stay competitive in a rapidly evolving market.

#### 4.3. Future Research Directions

Based on the background research of big data analytics and the proposed methodology and the model for supply chain BDA, the following main research directions can be defined:

- Real-time data integration—Investigate methods for integrating real-time data from diverse sources (e.g., IoT devices, social media, market trends) into supply chain management systems and assess the impact of real-time data on supply chain visibility and responsiveness;
- Predictive analytics and machine learning—Explore the use of predictive analytics and machine learning models to forecast supply chain disruptions and demand fluctuations and evaluate the effectiveness of these models in improving supply chain planning and risk management;
- AI applications—Examine the potential of generative AI for creating synthetic data to enhance supply chain simulations and scenario planning. Investigate how generative AI can be used to optimize supply chain designs and processes by generating innovative solutions and strategies. Assess the potential of LLMs (Large Language Models) in improving supply chain collaboration by facilitating better understanding and interpretation of complex data;
- Scalability and performance optimization—Examine the scalability of BI and BDA solutions in handling large datasets and complex supply chain networks and develop strategies for optimizing the performance of these systems to ensure efficient data processing and analysis;
- Data quality and governance—Address challenges related to data quality and governance in big data environments and propose frameworks for maintaining high standards of data integrity and security in supply chain operations;
- Cost-benefit analysis—Conduct a comprehensive cost-benefit analysis of implementing advanced BI and BDA tools in supply chain management. Assess the return on investment (ROI) for organizations, considering both direct and indirect benefits;
- Sustainability and ethical considerations: Explore how BI and BDA can support sustainable supply chain practices and contribute to environmental and social goals. Address ethical considerations related to data privacy and the use of AI in supply chain management.

## 5. Conclusions

Due to globalization and other market factors, supply chains are one of the key business systems. They not only need to be economically efficient but also need to be sustainable and improve the social and environmental impact. This creates many SCM challenges that can only be resolved with more intelligent decision making. This involves incorporating advanced big data analytical systems capable of integrating, storing, processing, and analyzing large and various volumes of data.

Over the last decade, there have been significant advancements in terms of variety, velocity, and volume of data created not only inside companies but also throughout supply chains. This brings various challenges related to business analytics, especially when it comes to infrastructure, platforms, software solutions, and development methodology. The present supply chain environment requires more collaborative, agile, flexible, and intelligent decision making.

Business intelligence technologies enable companies to achieve new and effective innovations that improve their supply chain processes, competitiveness, and sustainability. As supply chains become more and more complex, successful management of these complex processes becomes dependent on big data systems. Supply chain companies face numerous challenges related to big data integration, storage, modeling, processing, analysis, and visualization.

One of the main contributions of this paper is the methodology for designing and implementing supply chain BDA solutions that comprise the whole analytical life cycle and the specifics of supply chain systems. It incorporates an iterative and incremental approach with clearly defined phases and tasks while providing flexibility for creating various types of analytical solutions.

The presented multi-tiered SCM big data model and the cloud-/based architecture enable the design of novel loosely-coupled composite business intelligence systems, which bring together various data sources, information systems, data models, and visualization artifacts within the unified BDA ecosystem.

Cloud-based architecture with various managed analytical services such as data gateways, common data service, data lake, big data warehouse, and machine learning models, as well as artifacts such as predefined calculations, KPIs, data flows and pipelines, reports, dashlets, and content packs, enable the faster design, development, and deployment of composite big data analytical solution, adapted to specific supply chain analytical requirements.

The illustrative big data solution for supplier quality analysis demonstrates the applicability and effectiveness of the proposed big data methodology and its architecture. It shows how various cloud-based analytical services can be designed, reused, and composed into a user-friendly BI portal with dashboards, drill-down capabilities, automated machine learning analytics, and self-service BI, where decision-makers can monitor supplier-related processes, gain real-time valuable insights, take proactive action, and ultimately improve sustainability. While the presented illustrative solution intends to show the capabilities and the potential of the methodology and the architecture, it is yet to be evaluated in a live supply chain environment.

The main characteristics and advantages of the proposed approach can be summarized as follows:

- **Adaptivity**—The system can be adjusted according to specific supply chain needs, integrated with existing systems, and extended with new services and tools;
- **Faster development**—It is possible to develop or compose solutions more easily by reusing existing assets (data sources, data models, queries, KPIs, reports, etc.);
- **Scalability and performance**—Cloud-based architecture enables elastic scalability, high performance, and close to real-time analytics, with in-memory data storage;

- Rich data modeling—The BI semantic model serves as an additional model layer over data sources and enables the design of data models, analytical models, calculations, business rules, KPIs, etc.;
- User-friendly and effective data exploration and visualization;
- Intelligent cognitive services for more automated decision making and optimization (i.e., NLP, bots, digital assistants, etc.);
- Information sharing and collaborative decision making.

The described analytical models, services, and tools demonstrate the applicability and effectiveness of the presented big data approach. This approach enables the efficient creation of feature-rich, flexible, high-performant, collaborative, and secure supply chain analytical systems. Through the adoption of the proposed BDA model, companies should have a more straightforward way of achieving more adaptive, effective, intelligent, collaborative, agile, and sustainable supply chains.

Future work and research will be carried out to enhance, extend, and evaluate the presented BDA methodology, architecture, and models. This involves extending the common data model with more entities, data warehouse BI models with additional cubes and KPIs based on the Green SCOR, as well as the creation of more BI artifacts such as data flows, reports, dashboards, and process-specific BI content packs. Related to advanced analytics, more machine learning algorithms and models that relate to specific supply chain processes and that provide additional types of insights can be designed. Also, efforts will be made to deploy and assess the proposed BDA methodology and the model in the live supply chain environment to obtain feedback information about effectiveness and usability.

**Author Contributions:** Conceptualization, N.S.; Methodology, N.S., M.R., Z.B., J.P. and A.G.; Software, N.S., M.R., Z.B., J.P. and A.G.; Validation, N.S., M.R., Z.B., J.P. and A.G.; Formal analysis, N.S., M.R., Z.B. and J.P.; Investigation, N.S., J.P. and A.G.; Resources, N.S.; Data curation, N.S.; Writing—original draft, N.S., M.R., Z.B., J.P. and A.G.; Writing—review & editing, N.S., M.R., Z.B., J.P. and A.G.; Visualization, N.S., J.P. and A.G.; Supervision, N.S.; Project administration, N.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia, and these results are parts of Grant No. 451-03-66/2024-03/200132 with the University of Kragujevac—Faculty of Technical Sciences Cacak.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. UN Global Compact Office; BSR. *Supply Chain Sustainability: A Practical Guide for Continuous Improvement*, 2nd ed.; UN Global Compact: New York, NY, USA, 2015. Available online: <https://www.unglobalcompact.org/library/205> (accessed on 10 September 2024).
2. Da Silva, B.M.; Ferreira, D.H.L.; Georges, M.R.R. Sustainable Practices in the Supply Chain under the UN Global Compact Perspective. *Int. J. Innov. Educ. Res* **2019**, *7*, 135–153. [CrossRef]
3. Villena, H.V.; Gioia, A.D. A More Sustainable Supply Chain. *Harv. Bus. Rev.* **2020**, *98*, 84–94.
4. Greenstein, A.S. Sustainability Starts with the Supply Chain. Available online: <https://www.industryweek.com/supply-chain/supplier-relationships/article/21963955/sustainability-starts-with-the-supply-chain> (accessed on 20 August 2024).
5. Stefanovic, N.; Milosevic, D. Developing Adaptive Business Intelligence Systems for Agile Supply Chain Analytics. In Proceedings of the 1st International Conference on E-Commerce, E-Business and E-Government (ICEEG'17), Turku, Finland, 14–16 June 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 45–50. [CrossRef]



6. Liu, H.; Lu, F.; Shi, B.; Hu, Y.; Li, M. Big data and supply chain resilience: Role of decision-making technology. *Manag. Decis.* **2023**, *61*, 2792–2808. [[CrossRef](#)]
7. Tamym, L.; Benyoucef, L.; Moh, A.N.S.; El Ouadghiri, M.D. A big data based architecture for collaborative networks: Supply chains mixed-network. *Comput. Commun.* **2021**, *175*, 102–111. [[CrossRef](#)]
8. Duggineni, S. Data Analytics in Modern Business Intelligence. *J. Mark. Supply Chain. Manag.* **2023**, *1*, 1–4. [[CrossRef](#)] [[PubMed](#)]
9. Supply Chain Today. Top Supply Chain Challenges and Priorities for 2025. Available online: <https://www.supplychaintoday.com/top-supply-chain-challenges-and-priorities> (accessed on 22 December 2024).
10. Zheldak, P. Big Data in Supply Chain: Real-World Use Cases and Success Stories. Acropolium. Available online: <https://acropolium.com/blog/big-data-in-supply-chain-real-world-use-cases-and-success-stories> (accessed on 22 December 2024).
11. Stefanovic, N. Big Data Analytics in Supply Chain Management. In *Encyclopedia of Organizational Knowledge, Administration, and Technology*; Mehdi Khosrow-Pour, D.B.A., Ed.; IGI Global Scientific Publishing: Hershey, PA, USA, 2021; pp. 2443–2457. [[CrossRef](#)]
12. Reynolds, J. Big Data Analytics Challenges and Solutions. Datameer. Available online: <https://www.datameer.com/blog/challenges-in-big-data-analytics-and-solutions-to-tackle-them> (accessed on 21 December 2024).
13. Likhadzed, V. Top 5 Big Data Challenges and How You Can Address Them. ITrex. Available online: <https://itrexgroup.com/blog/5-big-data-challenges> (accessed on 21 December 2024).
14. Sethi, B.K. Sustainability Data Management: A Catalyst for Addressing Change. Kellton. Available online: <https://www.kellton.com/kellton-tech-blog/sustainability-data-management> (accessed on 20 December 2024).
15. Chen, D.Q.; Preston, D.S.; Swink, M. How the Use of Big Data Analytics Affects Value Creation in Supply Chain Management. *J. Manag. Inf. Syst.* **2016**, *32*, 4–39. [[CrossRef](#)]
16. Schoenherr, T.; Speier-Pero, C. Data Science, Predictive Analytics, and Big Data in Supply Chain Management: Current State and Future Potential. *J. Bus. Logist.* **2015**, *36*, 120–132. [[CrossRef](#)]
17. Zangara, G.; Filice, L. Innovating the management of supply chains for social sustainability: From the state of the art to an integrated framework. *Eur. J. Innov. Manag.* **2024**, *27*, 360–383. [[CrossRef](#)]
18. Kumar, K.; Kumar, A. Application of Optimization Models in Sustainable Supply Chain Management: A Systematic Review Based on PRISMA Guidelines. *Process Integr. Optim. Sustain.* **2024**, 1–28. [[CrossRef](#)]
19. Adewusi, A.O.; Okoli, U.I.; Adaga, E.; Olorunsogo, T.; Asuzu, O.F.; Daraojimba, D.O. Business intelligence in the era of big data: A review of analytical tools and competitive advantage. *Comput. Sci. IT Res. J.* **2024**, *5*, 415–431. [[CrossRef](#)]
20. Halper, F.; Krish, K. TDWI Big Data Maturity Model Guide: Interpreting Your Assessment Score. TDWI Research. 2014. Available online: <https://tdwi.org/whitepapers/2013/10/tdwi-big-data-maturity-model-guide.aspx> (accessed on 10 September 2024).
21. Ji, G.; Hong, W. Research on the Manufacturer’s Strategies under Different Supply Interruption Risk Based on Supply Chain Resilience. *Sustainability* **2024**, *16*, 874. [[CrossRef](#)]
22. Mageto, J. Big Data Analytics in Sustainable Supply Chain Management: A Focus on Manufacturing Supply Chains. *Sustainability* **2021**, *13*, 7101. [[CrossRef](#)]
23. Chalmeta, R.; Barqueros-Muñoz, J.-E. Using Big Data for Sustainability in Supply Chain Management. *Sustainability* **2021**, *13*, 7004. [[CrossRef](#)]
24. Accenture. Big Data Analytics in Supply Chain: Hype or Here to Stay? Available online: <https://iehost.net/pdf/Accenture-Global-Operations-Megatrends-Study-Big-Data-Analytics.pdf> (accessed on 12 September 2024).
25. Rowe, S.; Pournader, M. How Big Data Is Shaping the Supply Chains of Tomorrow. KPMG. Available online: <https://assets.kpmg.com/content/dam/kpmg/au/pdf/2017/big-data-analytics-supply-chain-performance.pdf> (accessed on 18 September 2024).
26. Stefanovic, N.; Milosevic, D. A Review of Advances in Supply Chain Intelligence. In *Advanced Methodologies and Technologies in Business Operations and Management*; Mehdi Khosrow-Pour, D.B.A., Ed.; IGI Global Scientific Publishing: Hershey, PA, USA, 2019; pp. 1211–1224.
27. Zhan, Y.; Hua, T.K. An analytic infrastructure for harvesting big data to enhance supply chain performance. *Eur. J. Oper. Res.* **2020**, *281*, 559–574. [[CrossRef](#)]
28. Renner, A.; Overcoming 5 Major Supply Chain Challenges with Big Data Analytics. IDG Communications. Available online: <https://www.computerworld.com/article/1640640/overcoming-5-major-supply-chain-challenges-with-big-data-analytics.html> (accessed on 20 September 2024).
29. Columbus, L. Ten Ways Big Data Is Revolutionizing Supply Chain Management. Available online: <https://www.forbes.com/sites/louiscolumbus/2015/07/13/ten-ways-big-data-is-revolutionizing-supply-chain-management/#3edf715f69f5> (accessed on 2 October 2024).
30. Gomez, J.M.F.; Top Challenges for Big Data in the Supply Chain Management Process. Advanced Fleet Management Consulting. Available online: <https://advancedfleetmanagementconsulting.com/eng/2016/07/13/1199/> (accessed on 5 October 2024).
31. Nguyen, T.; Zhou, L.; Spiegler, V.; Ieromonachou, P.; Lin, Y. Big data analytics in supply chain management: A state-of-the-art literature review. *Comput. Oper. Res.* **2018**, *98*, 254–264. [[CrossRef](#)]

32. Wamba, S.F.; Akter, S. Big Data Analytics for Supply Chain Management: A Literature Review and Research Agenda. In *Enterprise and Organizational Modeling and Simulation; Lecture Notes in Business Information Processing*; Barjis, J., Pergl, R., Babkin, E., Eds.; Springer: Cham, Switzerland, 2015; Volume 231, pp. 61–72. [[CrossRef](#)]
33. Waller, M.A.; Fawcett, S.E. Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management. *J. Bus. Logist.* **2013**, *34*, 77–84. [[CrossRef](#)]
34. Roßmann, B.; Canzaniello, A.; von der Gracht, H.; Hartmann, E. The future and social impact of Big Data Analytics in Supply Chain Management: Results from a Delphi study. *Technol. Forecast. Soc. Chang.* **2020**, *130*, 135–149. [[CrossRef](#)]
35. Niu, T.; Zhang, H.; Yan, X.; Miao, Q. Intricate Supply Chain Demand Forecasting Based on Graph Convolution Network. *Sustainability* **2024**, *16*, 9608. [[CrossRef](#)]
36. Ali, S.M.; Rahman, A.U.; Kabir, G.; Paul, S.K. Artificial Intelligence Approach to Predict Supply Chain Performance: Implications for Sustainability. *Sustainability* **2024**, *16*, 2373. [[CrossRef](#)]
37. Arunachalam, D.; Kumar, N.; Kawalek, J.P. Understanding big data analytics capabilities in supply chain management: Unravelling the issues, challenges and implications for practice. *Transp. Res. Part E Logist. Transp. Rev.* **2017**, *114*, 416–436. [[CrossRef](#)]
38. Cohen, A.M. Inventory Management in the Age of Big Data. *Harv. Bus. Rev.* **2015**, *94*, 1–5.
39. Renner, A. 5 Data-Driven Supply Chain Challenges—And What You Can Do About Them. Informatica. Available online: <https://www.informatica.com/blogs/5-data-driven-supply-chain-challenges-and-what-you-can-do-about-them.html> (accessed on 9 September 2024).
40. Gunasekaran, A.; Papadopoulos, T.; Dubey, R.; Wamba, S.F.; Childe, S.J.; Hazen, B.; Akter, S. Big data and predictive analytics for supply chain and organizational performance. *J. Bus. Res.* **2017**, *70*, 308–317. [[CrossRef](#)]
41. Ellis, S. Unlocking Big Data Cost Efficiency. The Knowledge Network. Available online: <https://www.theknowledgeacademy.com/blog/big-data-cost/> (accessed on 20 December 2024).
42. Kacyor, C. The 50 Best Supply Chain Analytics Tools and Software. Camcode. 2024. Available online: <https://www.camcode.com/blog/top-supply-chain-analytics/> (accessed on 21 December 2024).
43. Mohanty, S.; Jagadeesh, M.; Srivatsa, H. *Big Data Imperatives—Enterprise Big Data Warehouse, BI Implementations and Analytics*; Apress: New York, NY, USA, 2013.
44. ObviEnce. Supplier Quality Dataset. 2015. Available online: <https://github.com/BasseyIsrael/Supplier-Quality-Analysis> (accessed on 20 December 2024).

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.