OmniaScience

Journal of Industrial Engineering and Management
JIEM, 2022 – 15(4): 614-636 – Online ISSN: 2013-0953 – Print ISSN: 2013-8423
https://doi.org/10.3926/jiem.3922

Sustainable Manufacturing in the Fourth Industrial Revolution: A Big Data Application Proposal in the Textile Industry

Gustavo Araque-González, Albeiro Suárez-Hernández, Mauricio Gómez-Vásquez, Juan Vélez-Uribe, Alexis Bernal-Avellaneda

Politécnico Grancolombiano Institución Universitaria (Colombia)

garaque@poligran.edu.co, asuarez@poligran.edu.co, mgomezva@poligran.edu.co, jpvelezu@poligran.edu.co, eabernal@poligran.edu.co

Received: February 2022
Accepted: July 2022

Abstract:

Purpose: Design an industrial production model with a focus on industry 4.0 (Big Data) and decision-making analysis for small and medium-sized enterprises (SMEs) in the clothing sector that allows improving procedures, jobs and related costs within the study organization

Develop a sustainable manufacturing proposal for the industrial textile sector with a focus on Big data (entry, transformation, data loading and analysis) in organizational decision making, in search of time and cost optimization and environmental impact mitigation related.

Design/methodology/approach: The present research, of an applied nature, raises a value proposition focused on the planning, design and structuring of an industrial model focused on Big Data, specifically in the apparel manufacturing sector for decision-making in a structured and automated way. with the methodological approach to follow: 1) Approach of production strategies oriented in Big Data for the textile sector; 2) Definition of the production model and configuration of the operational system; 3) Data science and industrial analysis, 4) Production model approach (Power BI) and 5) model validation.

Methodological design of the investigation. 1) Presentation of the case study, where the current situational analysis of the company is carried out, formulation of the problem and proposal of solution for the set of data analyzed; 2) Presentation of a solution proposal focused on Big Data, on the identification of the industrial ecosystem and integration with the company's information systems, as well as the solution approach in the study and science of data in real time; 3) Presentation of the Model proposal for SQL structured databases in the loading, transformation and loading of important information for this study; 4) Information processing, in the edition of data in the M language of Power BI software, construction and elaboration of the model; 5) Presentation of the related databases, in the integration with the foreign key of the Master table and the transactional Tables; 6) Data analysis and presentation of the Dashboard, in the design, construction and analysis of the related study variables, as well as the approach of solution scenarios in the correct organizational decision making

Findings: The results obtained show an improvement in operational efficiency from the value-added proposal

Research limitations/implications: Currently, the number of studies applying Big Data technology for organizations in the textile and manufacturing sector in organizational decision making are limited. If analyzed from the local scene, there are few cases of Big Data implementation in the textile sector, as a consequence of the lack of projects and financing of value propositions. Another limiting factor in this research is the absence of digital information of high relevance for study and analysis, which leads to longer times in data entry and placement in information systems in real time. Finally, there is no data organizational culture, where there are processes and/or procedures for data registration and its transformation into clean data.

Originality/value: This research integrates, as well as the correct organizational decision making

For the verification of originality, the project search and systematic review of literature in the main online search engines are carried out for this research; In addition, the percentages of coincidence with online reviewers such as turnitin and plag.es are reviewed in the transparency of this study project.

Keywords: big data, industry 4.0, textile industry, sustainable manufacturing, data science, power BI

To cite this article:

Araque-González, G., Suárez-Hernández, A., Gómez-Vásquez, M., Vélez-Uribe, J., & Bernal-Avellaneda, A. (2022). Sustainable manufacturing in the fourth industrial revolution: A big data application proposal in the textile industry. *Journal of Industrial Engineering and Management*, 15(4), 614-636. https://doi.org/10.3926/jiem.3922

1. Introduction

The era of digitization, industrial transformation and the rise of new and innovative technologies have opened up the development of new approaches and production strategies within organizations. Production models currently seek a synchrony between the world of physical operational transformation and its interrelation with the digital world and analysis of the productive behavior of the value chain of each of the organizations (Zhong, Lan, Xu, Dai & Huang, 2016). The result: transformation and arrangement of finished products and the generation of large amounts of data, because of each of the human-machine (H-M) and machine-machine (M-M) interactions.

The generation of high volumes of data as a result of the operational processes of organizations is the main input for the analysis of behavior and improvement of industrial processes. In relation to the previous requirements, a new era of technological transformation known as industry 4.0 is presented, which is described by (Vaidya, Ambad & Bhosle, 2018; Rojko, 2017) as the "smart industry" focused on a new era of digitization and the efficient use of technological tools for the analysis, implementation, control and improvement of each of the customer's requirements, as well as the personalization and treatment of the industrial behavior of each of the orders. Big Data is presented as one of these technological tools, which seeks as an objective the incessant work of administration, analysis and proposal of value based on the resulting information. (Hernandez, Duque & Moreno, 2017; Lu, Li, Zhang & Yang, 2017).

Recent research in Wan, Tang, Li, Wang, Liu, Abbas et al. (2017), Y. Lu and Xu (2019) and Majeed, Zhang, Ren, Lv, Peng, Waqar et al. (2021) indicate that industrial automation and application of Big Data as a science and study of Operations works on setting up production processing systems and make-on-demand service application. Some of the promising scenarios focus on applications on manufacturing equipment driven by Cyber Physical Systems (CPS) that allow and assist in industrial automation in the "intelligent" configuration of industrial machinery and equipment to the dynamism of demand (Guo, Zhang, Zhao & Song, 2020; Lins & Oliveira, 2020). This research is framed in this development scenario, given the production needs and increased demands of the internal and external market of the organization.

One of the challenges of the industrial sector is the study of the behavior of the operational processes, information and resulting data in the product life cycle. In accordance with the aforementioned, this applied research seeks

Design an industrial production model with a focus on industry 4.0 (Big Data) and decision-making analysis for small and medium-sized enterprises (SMEs) in the clothing sector that allows improving procedures, jobs and related costs within the study organization.

To achieve the objectives set out above, this research focuses on the area of fabric dyeing operations in a study manufacturing company, developed in 5 main phases: 1) Construction, search and collection of information, where they are raised current production strategies, represented in flow diagrams and process information systems available, data analysis, operational capacity, industrial design and correlational analysis of the variables involved from a first situational diagnosis; 2) Definition of the production model, the configuration of the operational system, parameters, relevant variables, necessary resources, prioritization, assumptions and diagram of the operational models; 3) Data Analysis, Data Collection Plan (Sampling Technique and Sample Size); definition of parameters of each production model; 4) Approach of production model, use of specialized software (Power BI) to carry out the conceptual model of production of each scenario, focused on analysis of massive data - Big Data; 5) Validation of the model, Verification of the conceptual model versus the assumptions of each production model. It is expected that, through the results obtained in this research, the configuration and standardization of an industrial production model and the development of an optimal solution algorithm can be opened up in the lean accounting measurements for the industrial processes of the case study and its projection. in the industrial sector (textile cluster).

2. Literature Review

Industry 4.0 belongs to the most important fourth industrial stage that has been verified from the industrial revolution started in the XVII century. An initial assumption on this fourth industrial revolution, taking into account that due to Information and Communication Technology this revolution has reached every organization equally, is for the companies to be proactively prepared for the transformation potential by defining in advance manufacturing models, operational processes and the most suitable goals to confront associated challenges (Pereira & Romero, 2017). From a technical perspective, the Industry 4.0 context can be described as "a larger digitalization and the next automation level, in addition to a larger fit out communication by the creation of a digital, holistic and flexible chain value" (Oesterreich & Teuteberg, 2016) Even though it is not a fully accepted term, "Industry 4.0" remains a 'generic' concept (Pereira & Romero, 2017), however, the "Industry 4.0" term is an evolving trend and has entailed a greater and greater interest both in the professional and academic communities (Liao, Deschamps, Loures & Ramos, 2017; Fatorachian & Kazemi, 2018). Industry 4.0 includes mainly Internet of Things (IoT), cloud computing and cognitive computing, and digital manufacturing and the cyberphysics systems which gather, transfer and give meaning to Big Data in order to develop smart industries and respond to the VUCA (volatility, uncertainty, complexity, and ambiguity) environments of the current economy. Industry 4.0 has been used in manufacturing and in automotive industry by companies as BMW and Jaguar Land Rover, also in the food industry by companies as Mondelez and Nestlé to improve their general operative efficiency (Belman-Lopez, Jiménez-García & Hernández-González, 2020).

While it's found in literature a countless business and government projects of the BigData application where its benefits are openly recognized, it is not a widely democratized technology yet and there is still way to go through, principally in small and medium companies over Latin America (Belman-Lopez et al., 2020; Montecinos, 2021; Baines, Ziaee-Bigdeli, Bustinza, Shi, Baldwin & Ridgway, 2017). Industry 4.0 and specifically its Big Data technology implies an innovation wave and there are potential opportunities for organizations and supply chains to innovate, create a competitive advantage and generate a new commercial value as of large data volumes, with wide formats variety and quick access (real time) (Gandomi & Haider, 2015). In Colombia, Medellin specifically, the Center for the fourth industrial revolution project has been developed, a government project with private support which seeks: quality knowledge access, reach unique expert perceptions, be seen as regional and global leaders, get along with avant-garde technology and innovation and get Colombia to be more attractive to companies and invertionists (Belman-Lopez et al., 2020).

One of the industries that has seen significant benefits in technology introductions as Big Data is the textile sector. Fast fashion has one of the most negative impacts on the environment due to water use per processed clothing kilo and the high energy requirements needed to elaborate raw material, dye it, process it and place it to disposition to customers(Osorio & Vivas, 2021). Lately, researchers and legislatives have discussed circular economy solutions and innovative business models with the goal to reach sustainable development objectives on its production and responsible consumption components. The circular economy solutions have caught the textile industries, legislatives, and academics attention as we progress towards a more connected world. Industry 4.0 technologies can accelerate the industrial transition towards circularity and improve the weather sustainability indicators of this industry. Digital technologies help data transfers in real time on the condition, availability, accessibility and material and product resources, and drive the transition towards the Circular Economy in textile industries and clothing (Happonen & Ghoreishi, 2022). Product design development, prototype creations and material recycling can be fulfilled with higher efficiency using technologies such as Big Data. An outstanding factor is the Big Data implementation projects incorporation in Latin American companies, belongs to the high direction role in the needed organizational changes achievement, requires high direction to be willing to invest in new technology, in workers training and real manufacturing processes change.

3. Sample and Methods

3.1. Industrial Textile Case: From Empiric System to Textile Industry 4.0

The textile industry is presented as one of the sectors that generates the highest emission of environmental pollutants in the world (Singh, Singh, Gupta & Singh, 2019). Coloring and finishing processes require large amounts of water, as well as associated chemicals, present in colorants, dispersants and leveling agents. In the fabric dyeing process, the absence of industrial process control results in a high level of dye waste, which is lost in wastewater and generates a negative environmental impact. Given the situation raised above, it is necessary to generate strategies for the protection of the environment, society and industry, as well as the development of new procedures and production philosophies. The Green production have generated great attention in the scientific and environmental community as a proposed solution (Schoeberl, Brik, Braun & Fuchs, 2005; Nadeem, Guyer, Keskinler & Dizge, 2019).

One of the green production strategies allows identifying the need to transform empirical processes into digital activities, for better planning and control. A correct approach is presented by (Tsai, 2018; Guizzi, Falcone & De Felice, 2019) as the digital factory, being one of the relevant strategies in the textile industry 4.0 in search of mitigating the environmental impact generated in textile processes. The use of technology linked in Cyber Physical Systems (CPS), such as, for example, the mapping, visualization and real-time control of textile processes through specialized software allows organizations to carry out a survey in the short term and can generate environmental control action plans. The simulation of industrial processes and the visualization of performance measurement indicators support the correct placement of the industrial infrastructure and at the same time the minimization of waste is achieved in the real physical world (Abreu, Alves & Moreira, 2017; Sobottka, Kamhuber, Henjes & Sihn, 2017).

For the implementation of the virtual factory, it is important to generate a analysis, structuring and processing of the output information of the industrial system, necessary for the migration from physical space to the digital world. According to Khan, Wu, Xu and Dou (2017), Qi and Tao (2018) and Modoni, Caldarola, Sacco and Terkaj (2019), one of the approaches and great exponents of Industry 4.0 is Big Data development, being the integration of multidisciplinary technologies (connectivity, computing, industrial machinery, cyber-physical systems, cloud manufacturing, etc.) in an integrated technological ecosystem that allows visualizing digital industrial behavior for correct business decision-making. To integrate this set of technologies, this research plans, develops and analyzes the correct migration of physical information systems to the digital operational world, using Big Data technology, based on a business case.

The industrial behavior of the present case study is shown in Figure 1. The main organizational areas of a textile company and the role of each of the departments are exposed: 1) reception of the fabric, where the fabric inputs are received at be processed and assign the batch certificate, which guarantees the traceability of the horizontal process; 2) fabric preparation, place where the fabric input is unwrapped and placed in a filleting machine to go to the closing process; 3) dyeing, process in which the water is prepared with heat exchangers whose energy base is coal, reception of the mixture (chemical) to be applied in the water as well as the specifications of the dyeing curve, preparation of the machinery (round or long, depending on the type of processing): insert the type of curve, associated water levels, introduction of the fabric and placement of additive, dye and salt and finally use of the conveyor bongos for their output; 4) formulation and programming: in the case of the first, the formulation of the production batches is carried out, identifying the quantities of chemical components to be placed on the fabrics; and for the second element, it is known as the allocation of the productive sequence in a determined period of time; 5) finishing the fabrics, once the fabric leaves the dyeing and reaches this process, in this stage the conditioning of the fabric is presented: touch, weight and elongation are some of the physical characteristics developed in this process. 6) Quality, department responsible for communication with the client and verification of agreed standards: tone, solidity, width, weight, performance, etc.; 7) logistics, process in which the chain of distribution and transport of orders to the customer is generated; 8) Maintenance area, operational department responsible for corrective and preventive maintenance activities of industrial machinery, as well as their management and traceability; 9) Financial administration, responsible for the accounting and economic structure of the company, as well as the administrative and managerial processes; 10) Laboratory, this being the support area where the developments of new shades are generated, made up of based on chemicals, pipetting with water and the use of additives in experimental tests.

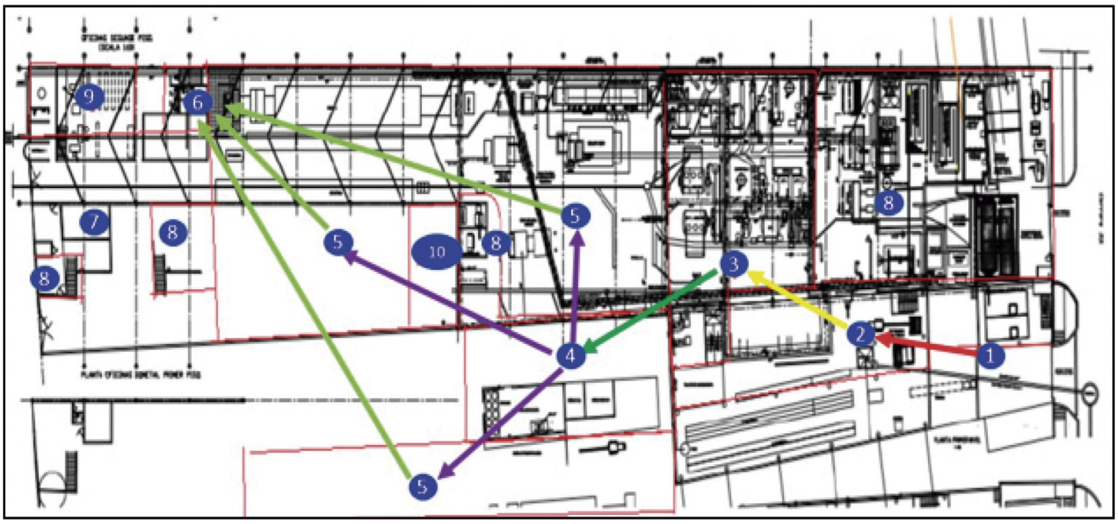

Figure 1. Physical Layout - case study

According to Figure 1, there is an interrelation of the operational departments through flows of materials, personnel and associated information. And it is on this last element where the development of this research is focused: at present, the company does not have a data centralization system, information protocols and standardization of operating language, so the decisions generated present a behavior empirical, not standardized and based on value judgments. To solve the above, this study develops a proposal for the application of information systems using Big Data technology for the textile industry 4.0, focused on the migration of information systems of physical operations and their migration towards a digital technological proposal. supported in the visualization, control and execution of the data from the Power BI software. This type of software, according to Jain (2017), Doko and Miskovski (2020) and Pinheiro (2020) is a business analysis tool that facilitates the process of ordering, structuring and data analysis, comprehensive visualization of processes in real time. Its main function is the transformation of non-coherent data sources into "intelligent" visual displays, based on Big data, that support efficient organizational decision-making.

3.2. Model Proposal for the Big Data Application in Textile Industry 4.0- Architecture and Overview

Big Data technology and its application in the professional field, according to Wang, Yang, Wang, Sherratt and Zhang (2020), Firouzi, Rahmani, Mankodiya, Badaroglu, Merrett, Wong et al. (2018) and Petrillo, Picariello, Santini, Scarciello and Sperlí (2020) can improve coordination and control systems social, productive efficiency, development and evolution of technological processes and industrial innovation. There are activities that cannot be approached with traditional methods; Therefore, the use of innovative technologies is essential in the correct implementation and action plan for organizational decision-making.

The industrial behavior of operations demands a correct optimization of costs, times and associated materials. According to Zhu, Yu, Wang, Ning and Tang (2019) and Wong, Zhou & Zhang (2019) this value strategy can be executed through Big Data technology, which seeks the correct treatment and study of massive data sources of industrial processes. The study of these distributed data sources requires storage and processing in appropriate technological infrastructure (hardware and software) associated with each of the operational processes in the production sequence, through the development of ecosystems for the integration of information systems and production analytics, also known as the Architecture and Industrial Big Data Overview.

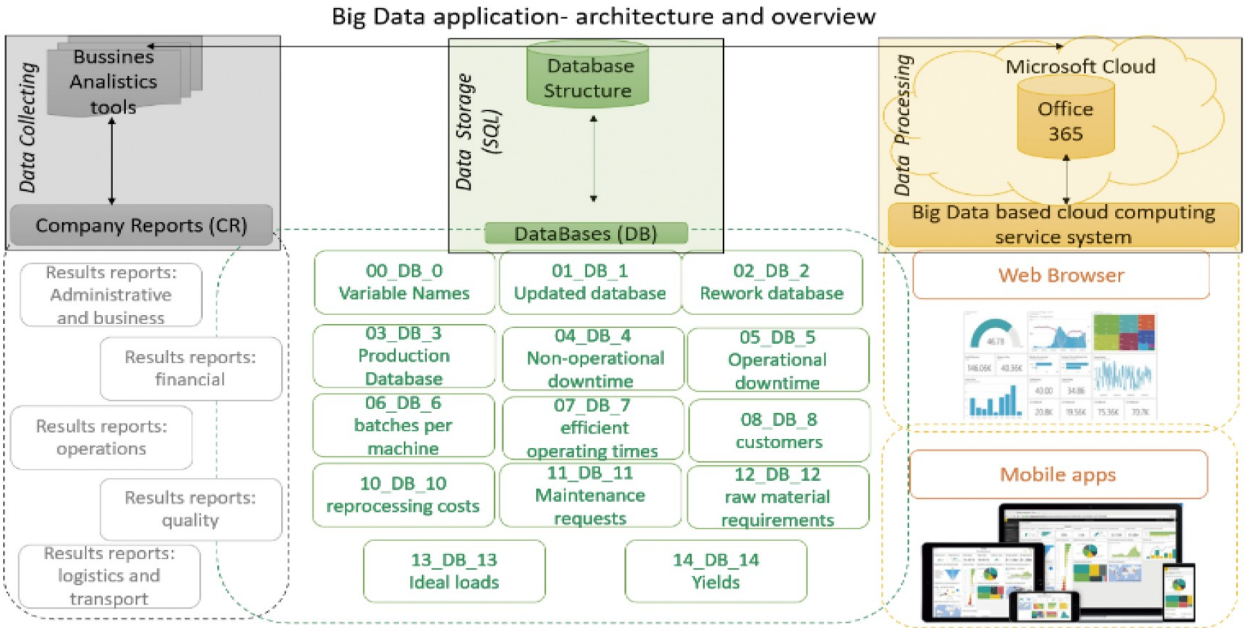

Figure 2. Big Data application-architecture and overview

Figure 2 illustrates the Big Data architecture and overview implemented for the textile industry 4.0 study. Each of the processes within the company (administration and business, financial, operations, quality, logistics and transportation) generate reports in real time (CR) in computer systems (data collection), which are represented in two main categories: 1) static data, those resulting from batches and related information present static storage (.xls); 2) dynamic flow data, being those obtained through metadata (.csv) and information outputs in systemic-industrial interactions. (Santoso & Yulia, 2017).

The collection of company data (CR) is not enough for the implementation in Big Data technology: the correct treatment, analysis and placement of the information associated with "structured" data categories is necessary and, in relation to Wang et al. (2020) and Saura, Herraez and Reyes-Menendez (2019), are known as those data architectures whose main characteristic is the presentation of distributed data systems SQL (Structured Query Language), in such a way that the presentation of information is Obtained in table format with rows and columns, resulting in Databases (DB). This SQL information allow for the third phase of the model, in data processing, to generate a specific category of information to be processed: Data Storage information, from data systems distributed in databases (DB) (Sun, Liu, Chen & Du, 2020). The company uses the Office 365 operating system to enter SQL data in the specific Power BI software, as a visualization software for the resulting information and associated variables. The type of Big Data technology used allows the generation of a service system based on cloud computing, where the reports and monitoring of the analyzed variables can be stored and analyzed in real time, once the SQL data has been inserted. With the download of Power BI software and considering the previous process, the hardware visualization is possible through personal computers, desktops and mobile devices, as well as its updating, monitoring and control.

3.3. Model Proposal Structuring - Structured Query Languages -SQL Databases

The language used in the databases of this study is SQL (Structured Query Language) managed in distributed databases (data storage), which has as its main characteristic its ease of access to information and interpretation, without the need to have prior knowledge of Query code (Nethravathi, Amitha, Saruka & Bharath, 2020; Wang, Liu, He & Wang, 2018). This type of language is characteristic for the correct reading and interpretation of queries in databases and information systems. The architecture and an overview of SQL databases is illustrated in Figure 3; It starts with the personal gateway (Power BI Personal Gateway) of the information associated with the industrial processes and services of the case study. In relation to the Big Data characteristics, the information is stored in file systems distributed in the Data Storage Server (SQL) with a data and registry extension (.MDF) (.LDF). This main characteristic allows the desired transactions to be registered and executed through Power BI Data Gateway, which functions as a gateway, link and address of local information, supported by the internal data security system. Next, the data source located in Power BI Data Gateway and Power BI Personal Gateway arrive at the Power BI dataset service, where this output information is collected and listed in its internal storage: each data set represents a single source of data. The dataset service is the main input for the development and implementation of the related applications in Power BI: 1) Reports, being the visual reports (tree maps, graphs, lines, maps, among others) that can be generated and shared (Microsoft cloud office 365). The Power BI reports service allows working under two modalities: reading view, for viewing with restricted access or it can be accessed through the editing view, in which the information contained can be accessed and modified according to user permissions registered with the company; 2) dashboards, an associated Power BI service whose main feature is the visualization of unique reports, usually created from the author Power user, in order to illustrate the work generated in the interaction of the distributed databases Data Storage Server (SQL) and study of variables. The great advantage of the dashboards service is that it allows these types of reports to be saved in real time, edited and shared with other users who have access to them. As an output, multiple users with access to Power BI services can view, edit and generate dataset developments in real time through the information generated through mobile or desktop devices.

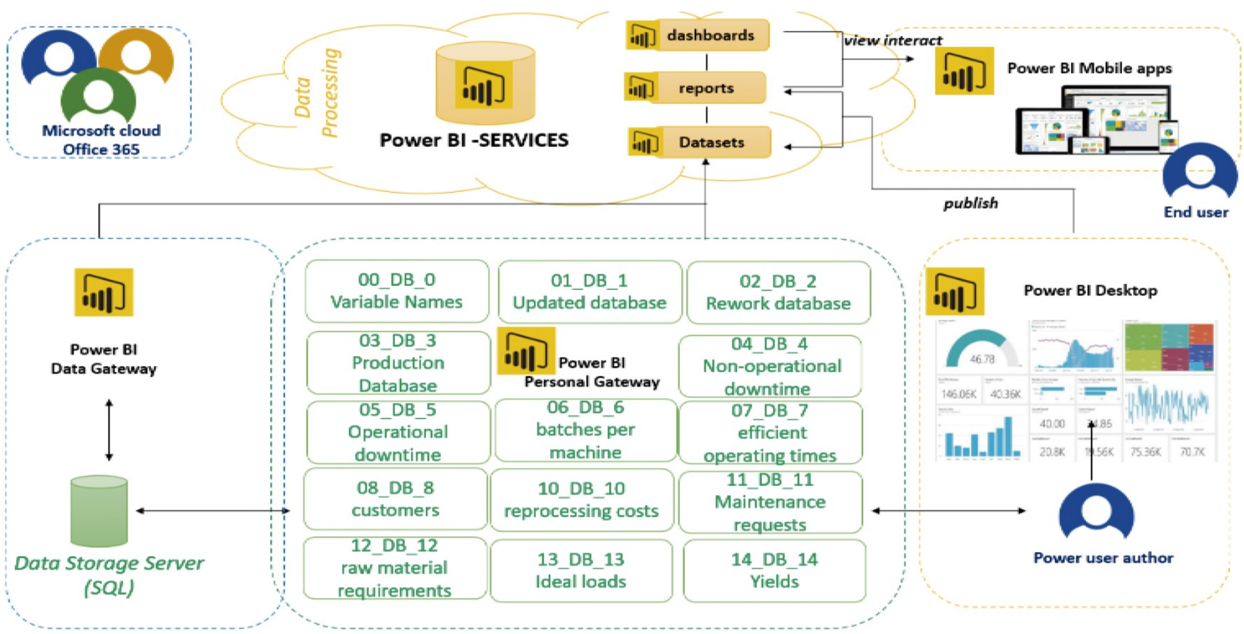

Figure 3. SQL Databases -architecture and overview

The applications proposed with the SQL data architecture, as well as the display and reporting panels, allow to control the storage, addressing and development of information systems in real time in an efficient, effective and effective way for all types of organizations. The migration of physical information to the digital environment (SQL), according to Gamero (2021) and Shao, Brodsky, Shin and Kim (2014) minimizes the environmental impact on the use of material resources related to paper consumption, energy use, water resources, among others. For the operational area, the correct placement and data structure of the present Big Data technology in the textile industry 4.0 allows the consolidation of data and presentation of compound reports, which facilitates its interpretation and adaptability to internal and external processes. (Petković, 2017) states that the monitoring and control of SQL databases must be carried out under a standardized report display pattern, which can be obtained with the correct use of metadata, migration and visualization, in relation to the specific needs of each operating department. The integration of information and its configuration from software with Big Data technology support rapid adaptability, sequence and traceability in industrial behavior: Banks (Bhalotia, Hulgeri, Nakhe, Chakrabarti & Sudarshan, 2002) allows generating advanced searches based on keywords obtained through of databases; SODA (Blunschi, Jossen, Kossman, Mori & Stockinger, 2012), generates SQL statements based on keywords, using the graphical pattern of the algorithm; The BOARDS system (Simos, Zivanovic & Leithner, 2019) supports the migration of information from a database without prior knowledge and / or training prerequisites in relation to programming or SQL. Some others, such as the BANKS models (Kleerekoper & Schofield, 2018), allow the interconnection of graphics in cloud computing interconnected with external links. The combination of navigability and determination of key words is part of this case study with the purpose of obtaining related information in distributed databases (SQL) and allows efficient access and use in information systems.

3.4. Data Processing - Preparation and Presentation

The presentation of the results obtained through the organizational databases has, as a characteristic, a high volume of output information in relation to the industrial processes under study. There are scenarios where "dirty data" can be generated, that is, information of low added value to be used within the real information to be analyzed and that directly impact the efficiency of the "real data" exposed in database systems. distributed data (SQL). Some of the causalities are exposed by (López-Porrero, 2011) through the following scenarios: 1) Data collision, characterized by the receipt of multiple information from various sources; 2) generation of multiple varieties of data, where results categorized into three main groups can be generated in the results of the database: structured, semi-structured or unstructured data; 3) syntactic anomalies, or also known lexical, format or domain errors and the lack of standardization of the information; 4) semantic anomalies, referring to contraindications of data values identified in tuples; 5) context anomalies, being the absence of data values in tuples due to lack of officialization of the information; 6) migration between information systems, a relationship in which a data transport process is generated, from one information system to another, generating errors in the information within this operation, among other error variables.

The process that is carried out for the purification and correct placement of the real data is exposed by Arzamasova, Schaler and Bohm (2018) and Giannakopoulou, Karpathiotakis, Gaidioz and Ailamaki (2017) as data cleaning. This type of operational characteristic implies the analysis in the error taxonomy and the necessary adjustments for the transformation of "dirty data" into useful information, through the development of a cleaning algorithm in the analyzed data set. In the present investigation, a cleaning and debugging algorithm in M language is used in Power Query Editor (Figure 4).

An example in the debugging of the present investigation is presented in Figure 4. The relation of the database 1_maestra (BD1 MAESTRA) relates the production batches for the orders of the study case and the dates in which the reception is made, generation of order entry form, operational dyeing processing period, dispatch of finished products and place of destination to customer. In the example illustration, the "origin" can be identified as the beginning, which relates the tracking and identification of the structured database type file (SQL) obtained from the company records and through the data editor (Power Query) it is carried out, in a structured and chronological way, the order in the treatment and purification of the information, according to the needs of integration of the data to be analyzed. Next, it is necessary for the present investigation through the data editor to generate the "promoted headers", which list as the main heading of the columns the titles of the variables that will be exposed and studied in the data analytics. As a third element, it is necessary to carry out a series of refinements, edits and adjustments in the configuration of the data, for its correct presentation (clean data): "Changed type", in the edition of the column type as a unit of measurement; "Replaced value", in the edition and adjustment of the necessary alphanumeric characteristics required by the data; "Aggregate conditional column", to generate labels known as mathematical conditionals that relate the information between each set of columns obtained through this SQL file.

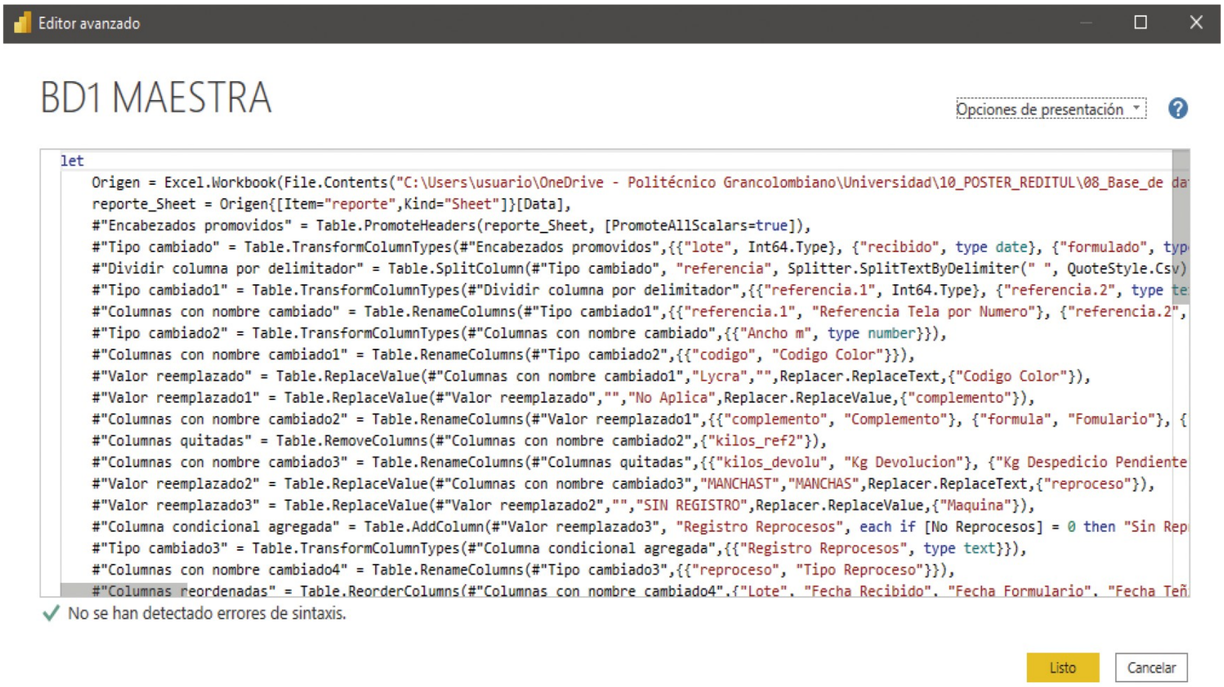

Figure 4. Database editing example in Power Query

3.5. Structuring the Model - Related Databases

The generation of "clean" databases allows the development of the integration and analysis in a coherent and real-time way of the information, according to the industrial behavior of the company. In accordance with (Aspin, 2016), for the correct analysis and integration of the databases, it is necessary to identify, link and develop a routing system through a "foreign key" whose characteristic is to be a common variable (column type - Power BI) for all databases that is the link between the entire set of information from the group of structured databases SQL to be analyzed (Figure 5).

For the correct operation of the "foreign key", two main types of databases must be identified within the model structure: The first is the master table database, whose characteristic is to be a static table with character information fixed in real time; the second is the transactional table database, in which the processing of information that occurs because of the company's operational behavior is recorded in real time. As a result of the process, each of these categories of databases must contain the "foreign key" to allow them to be integrated and analyzed in real time, transforming the system into related databases.

In this research, the two types of databases are identified: The "Master Table" database is known as: "01_DB_1 Updated database" with the internal information of ID Lot number, receipt date, formulation date, date of operation, client, reference, order, description, remission, rolls, code, chemical formula, type of process, input kilos, associated machine. In the case of "Transactional Tables" databases, the following databases are identified: 02_DB_2 Rework database; 03_DB_3 Production Database; 04_DB_4 Non-operational downtime; 05_DB_5 Operational downtime; 06_DB_6 batches per machine; 07_DB_7 efficient operating times; 08_DB_8 customers; 10_DB_10 reprocessing costs; 11_DB_11 Maintenance requests; 12_DB_12 raw material requirements; 13_DB_13 Ideal loads; 14_DB_14 Yields (Figure 3). In each of the transactional databases there is the "foreign key": batch number ID, which allows linking to the master database and thus working in an integral way the information in the data analytics to be studied.

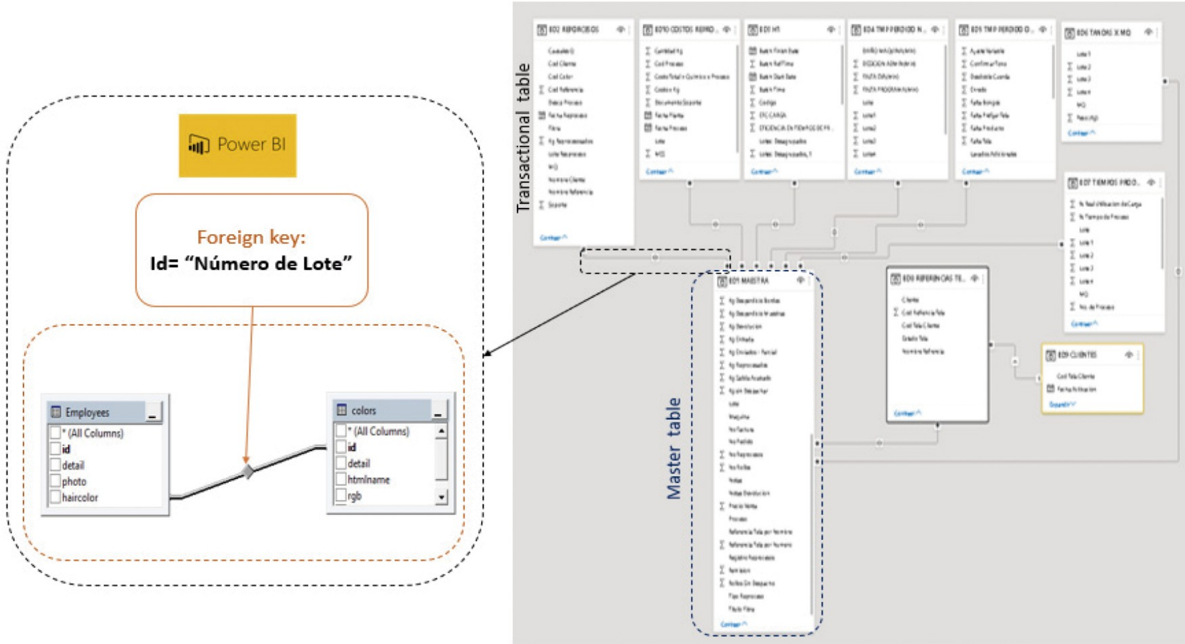

Figure 5. Related databases example in Power BI

3.6. Data Analytics and Dashboard Presentation

The data analytics process for research links the databases through the model proposed in Figure 5. This data integration (Master Table and transactional tables) allows generating reports of the desired variables and exposed in the databases. in real time, according to the information needs. With the use of the foreign key “ID batch number” it is possible to connect the information of the Master Table “01_DB_1 Updated database” with the information of “02_DB_2 Rework database”, as illustrated in the following example:

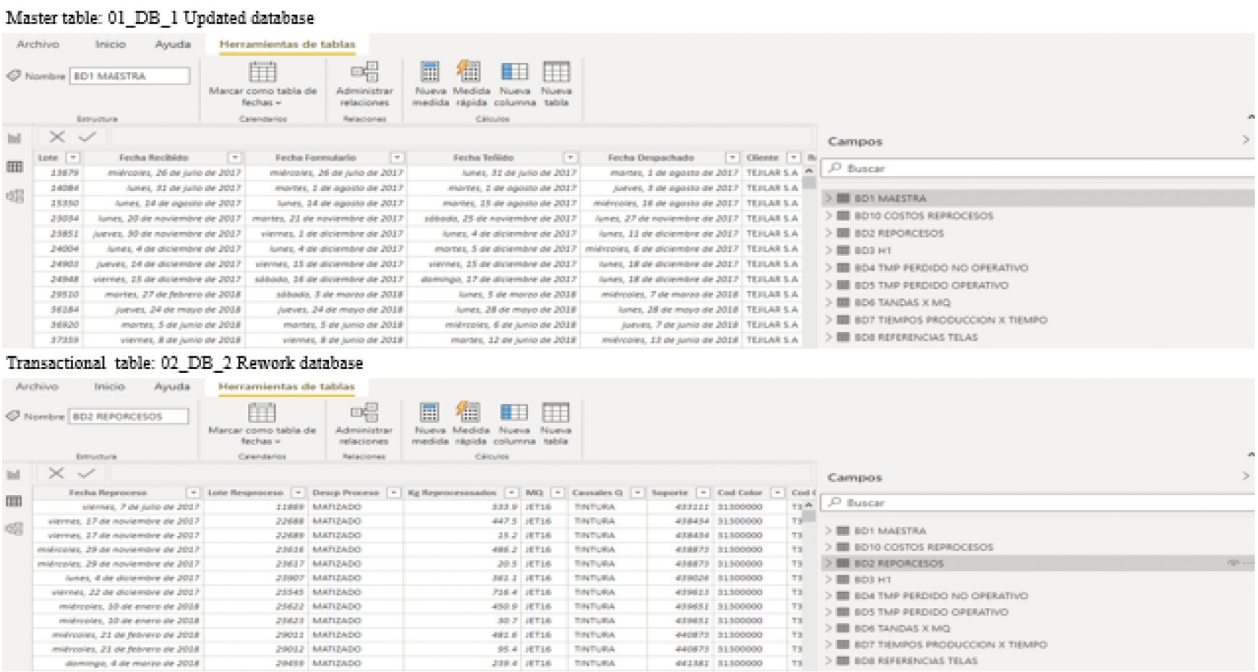

Figure 6. Matrix databases example in Power BI

Figure 6 illustrates an example of the relationship between the databases and their visualization through the Power BI software, as well as the foreign key that unites them called “batch number ID”. This binding feature allows the information obtained to be shared between each relationship set, such as the one shown in this example with the relationship type: one-to-many. In this way, it is possible to share the information of each of the variables displayed in columns in the Master Table "01_DB_1 Updated database" (Lot, Received date, Form date, Dyed date, Dispatched date, Customer, Fabric description, Referral, number of rolls, width, form, process, state, input kg, kg to be processed, waste kg, dispatched kg, output kg, invoice number, lot assignment date, days of fabric in the company, code color, among other variables) with the information displayed in the transactional table "02_DB_2 Rework database" (Batch, Reprocessing date, reprocessing description, reprocessed kg, cause, support number, color code, customer name, reference code , reference name, fiber, among other variables), in such a way that multiple and different scenario analyzes can be generated, relating and linking variables from one database to another. The process described above is known as data analytics, using three main groups of data analysis in this research.

- *Descriptive analysis*: The development of analysis models under the descriptive modality, according to (Patriarca, Chatzimichailidou, Karanikas & Di Gravio, 2022), seeks to extract information from existing records (historical data) to search for behaviors in the information systems analyzed. This type of behavior can be represented in reasons for success or failure in multi-scenario interaction.
- *Predictive Analysis*: This type of analysis, according to (Siegel, 2013) has as its characteristic the prediction of future behavior of systems analyzed and incorporated technologically. The operation of this category of analysis requires the extraction of information from data and its use for the analysis of trends and patterns of past, present or future behavior. This type of operation is what is known technologically as "data mining". (Gawusu, Mensah & Das, 2022).
- *Prescriptive Analysis*: A comprehensive and investigative analysis is presented under this prescriptive modality. According to (Rousopoulou, Vafeiadis, Nizamis, Iakovidis, Samaras, Kirtsoglou et al., 2022), prescriptive analysis not only seeks the characterization of past, present, and future systemic behaviors, it also links, through inferential analytics, knowledge of the causes of behavior. The main characteristic of this type of analysis is the technological synergy of data, business guidelines and mathematical models.

As a result of the present analysis of information obtained in Power BI and considering the analysis presented above, the integration of Master Databases with transactional databases allows the visualization of information in the form of figures, tables, timelines, among other tools, interactively in real time, as can be seen in Figure 7. This type of visualization is known as the Power BI "dashborad", which represents in real time and interactively the information obtained from the databases of the analyzed systems, as well as the interaction of the variables in the different types of analysis exposed.

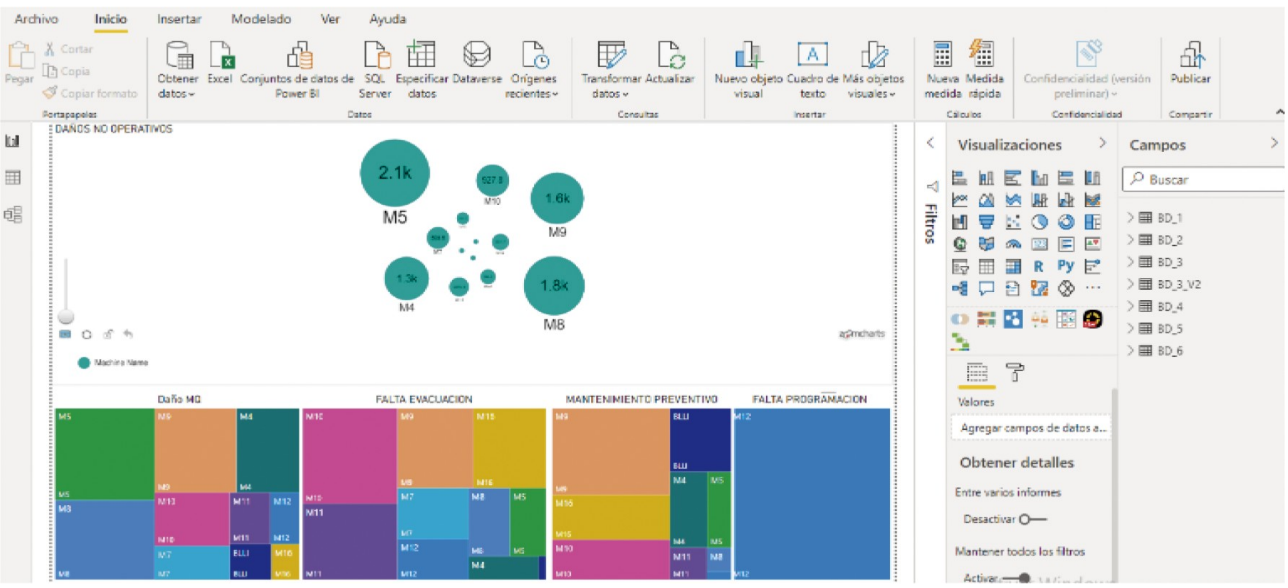

Figure 7. Matrix databases example in Power BI

4. Experimental Study and Analysis

4.1. Description of Business Requirements and Related Databases

As a result of the present analysis of information obtained in Power BI and considering the analysis presented above, the integration of Master Databases with transactional databases allows the visualization of information in the form of figures, tables, timelines, among other tools, interactively in real time, as can be seen in Figure 7. This type of visualization is known as the Power BI "dashborad", which represents in real time and interactively the information obtained from the databases of the analyzed systems, as well as the interaction of the variables in the different types of analysis exposed.

Information system	Process	Main variables managed
Visual Fox Pro	Production	Daily production, reprocessing, billing, batches, kilos, raw material, fiber code, client, final conditions.
Contai	Accounting	Suppliers, accounts payable, accounts receivable, customers, invoices.
Nominai	Payroll	Personnel, identification cards, social security, compensation fund, hours worked, overtime, payroll, disabilities.
Sedo máster	Dry-cleaner	Dyeing machines, process times, process stoppages, production planning, batches, bath ratio, kg.
Color máster	Chemical lab	LAB coordinates, shade reading, shade review, shade development, formula control.

Table 1. Organizational Information Systems

4.2. Related Database Information and Elaboration of the ETL Model (Extraction, Transformation and Loading)

4.2.1. Master Table 01_DB_1 Updated Database

Contains customer information, as well as the order reference and traceability through the system, with the following attributes: Production batch, customer code, number of rolls of fabric to be processed, chemical formula to be applied on the fabric, state, input kilos, production loss, output kilos, machine that performs the process, number of processes, price, invoice.

4.2.2. Transactional Table 02_DB_2 Rework Database

It contains the data that describes the reprocesses presented within the system, as well as the amount of waste generated within the process, once the number of kilograms of fabric entering the system is analyzed, processing of these and the number of kilograms of output, through the following attributes: Production batch, reprocessing date, applied process, kilos, machine, cause, color code, customer, reference, and fiber type.

4.2.3. Transactional Table 03_DB_3 Production Database

It contains the data that relates the total production of the system, as well as the characterization of the processes and the total production times for the customer batches (Batch time), with the following attributes: Production batch, machine, order number, batch grouping, processing code, type of process, Batch start date, Batch finish date, Batch time efficiency, load efficiency, amount of water involved in the process.

4.2.4. Transactional Table 04_DB_4 Non-Operational Downtime Database

It contains the information that relates the non-operational operational times and loss generators within the industrial system, through the following attributes: Production lot, machine, type of programming fault, type of EVA fault (Energy, Steam, Water), number of minutes in preventive maintenance, number of minutes in corrective maintenance, type of machine damage, number of minutes in human resource decision making.

4.2.5. Transactional Table 05_DB_5 Operational Downtime

It contains the information that relates the operational operating times and loss generators within the industrial system, through the following attributes: Production batch, machine, cause code: unsealed lycra, cause code: busy operator, cause code Cause Code: No Fabric, Cause Code: Other, Cause Code: Confirm Shade, Cause Code: Batch Wrong, Cause Code: Additional Washes, Cause Code: Tangle, Cause Code: Variable Setting.

4.2.6. Transactional Table 06_DB_6 Batches per Machine

It contains the data that relates the number of times (frequency) that a machine is used, as well as the occupancy and efficiency levels, based on the following attributes: Production batch, machine, weight in Kg, frequency in times, effective frequency, ineffective frequency, number of washes.

4.2.7. Transactional Table 07_DB_7 Efficient Operating Times

It contains the data that relates the efficiency of the operational times of each piece of equipment and associated batches, as well as occupancy levels and associated times, related to the following attributes: Production Batch, machine, process code, weight, Batch stop time, Batch ref time, Batch run time, Batch idle time, machine utilization, efficiency in process times.

4.2.8. Transactional Table 08_DB_8 Customers

Within this database, the main information of the personal data of the clients is related, as well as the code of each one of them within the organization, contained in the following attributes: Production lot, ID, order reference, name, address, city, telephone, email.

4.2.9. Transactional Table 10_DB_10 Reprocessing Costs

Within this database, the costs associated with the reprocesses generated in the industrial system of the study company are listed, accompanied by the following information: Production batch, invoice, process, product cost per Kg, kilos reprocessed, date, total cost of reprocessing.

4.2.10. Transactional Table 11_DB_11 Maintenance Requests

Within this database, the main information about the requests for preventive and corrective maintenance generated in each of the industrial equipment is listed, including the following attributes: Production lot, order number, machine, report date, date of repair, issue, responsible person who reports, response to maintenance request, solution code, type of solution.

4.2.11. Transactional Table 12_DB_12 Raw Material Requirements

Within this database, the main information about the demand for raw material within the production process is listed. The consumption of each of the system's inputs is identified, with the following attributes: Production lot, date, purchase reason, support, quantity, color code, purchase date, receive, weigher, cost.

4.2.12. Transactional Table 13_DB_13 Ideal Loads

Within this database, the ideal system load table is listed, that is, the amount in kilograms that must be introduced into each of the pieces of equipment, considering the variables associated with raw materials, supplies, equipment, and tools. associates, all of the above is identified within the following attributes: Production lot, type of fabric, machine, cords, minimum process in kg, load intervals, % yield.

4.2.13. Transactional Table 14_DB_14 Yields

Within this database, the performance that can be identified for each meter of fabric involved in the system is related, in relation to the chemicals used for its transformation, which contains the following attributes: Production batch, ID, type of fabric, fabric id, fabric yield.

With the definition of the Master database and transactional databases, the logical model of related databases is made, illustrated in Figure 8.

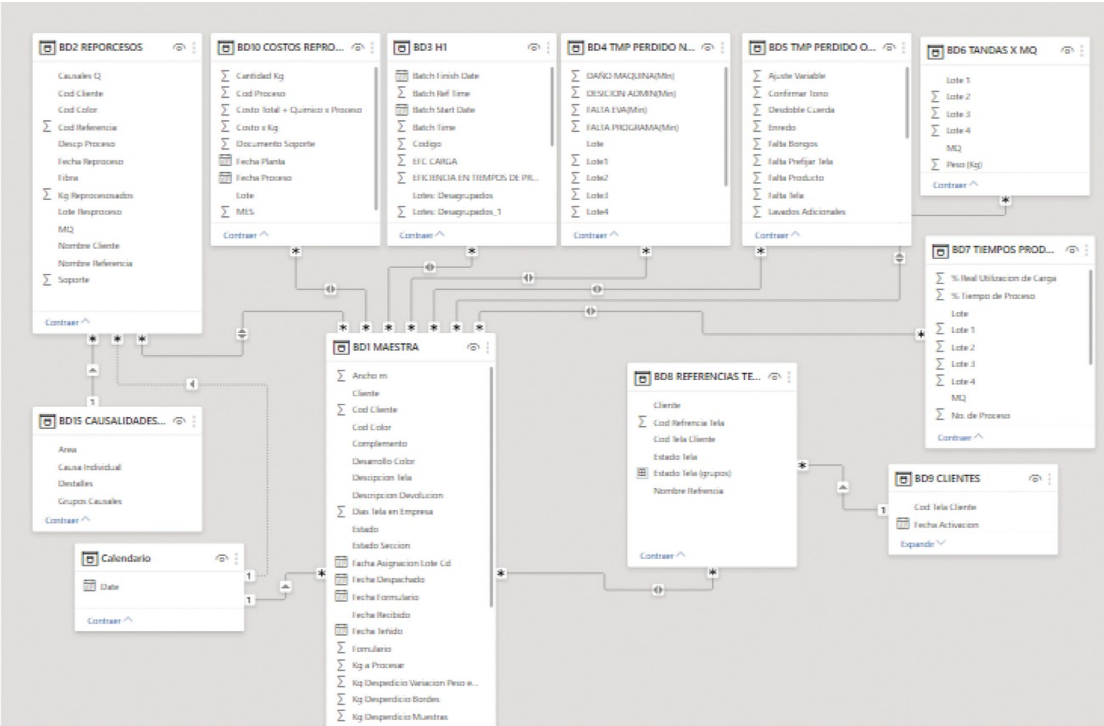

Figure 8. Logical model of related databases in Power BI

4.3. Results and Analysis

4.3.1. Analysis of Kilograms Reprocessed in the Industrial System

As a first step in the industrial requirements of the case study, the operational flow system and the results of kilograms reprocessed at the outlet are analyzed. In the described system, the production behavior is generated for four analysis periods (2017, 2018, 2019 and 2020), a control process generated since the creation of the organization. From the present investigation, the production department is analyzed, in relation to the percentage of participation of the clients, kilograms of input and output of the system, as well as the kilograms reprocessed from the industrial transformation process (Figure 9).

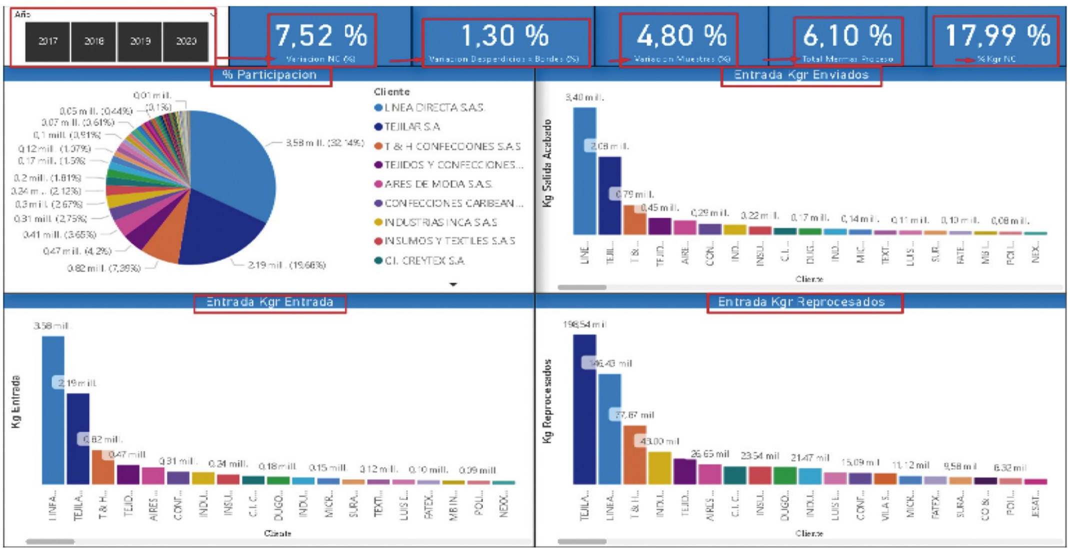

Figure 9. dashboard of kilograms reprocessed in the industrial system

In relation to the results obtained and in accordance with the requirements of operational control, the present investigation allows to identify exactly the percentage of participation of each one of the clients of the company (% Participation) and the methods of operational control of kilograms of entrance to the system as raw material or inputs (input kg shipped) as well as the kilograms output and sent to the customer's destination (input kg shipped), all of the above generating within the process the "wastes" or "non-conforming products" that are reprocessed to meet all the demand requested by customers. This type of behavior is important for the organization to establish the level of importance in sales generated by each type of customer and if, at a mathematical and argumentative level, the number of reprocessed kilograms is justified in the final profit margin that the company wishes to establish. According to the analysis exposed above, the analysis can be carried out for the client number 2 - dark blue in Figure 9. This type of customer represents 19.68% of the company's total sales for the analyzed period, occupying second place in all customers; however, it sends 2.19 million tons of fabric for industrial processing, 2.06 million tons of finished products are obtained and as a result, it generates for the organization, according to the requested orders, 198 thousand tons of associated "non-conforming product". among other types of resulting waste. This type of behavior allows the company to analyze whether it really justifies, according to its profit margin, having a client who, despite not being the company's most relevant client, leads the generation of non-conforming product within the operational processes of the company, and make future decisions in the cost-benefit analysis.

4.3.2. Causality Analysis of Reprocessed Kilograms (Quality Defects)

A second relevant element for this research prescriptively analyzes the causal behavior of the number of kilograms reprocessed in the presented system. This session analyzes, inferentially, what is the main causality that is generating non-conforming products, as well as the equipment involved in production. For this investigation, the kilograms processed per machine (Kg x MQ), kilograms of input to the system as raw material or inputs (input kg sent), percentage of cause of identified reprocesses and machinery involved (kg Reprocessed x causal MQ) are considered for this investigation, as well as the % of fibers (raw material) affected (Kg reprocessed x fiber x MQ). The results are presented in Figure 10.

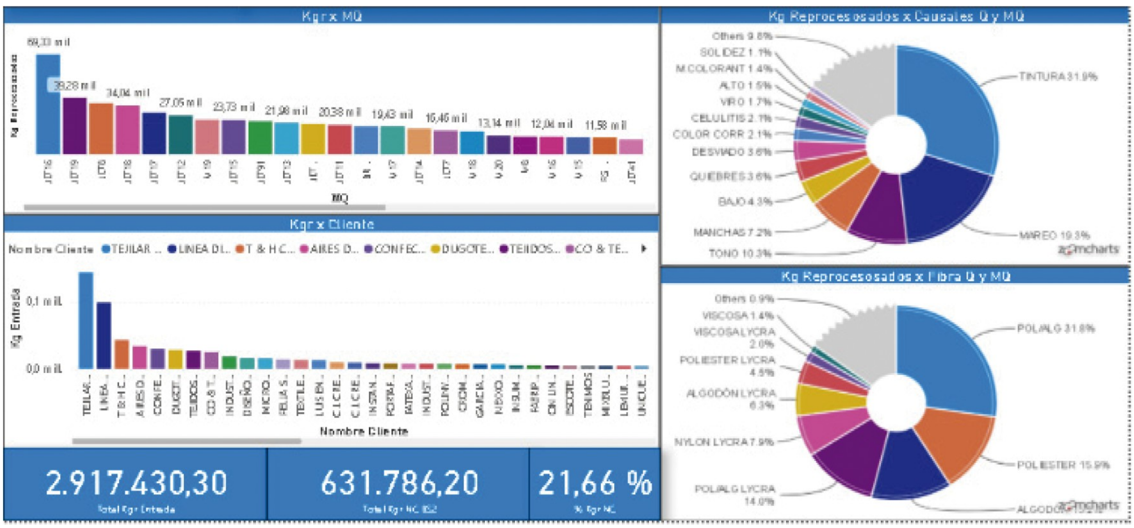

Figure 10. dashboard of kilograms reprocessed in the industrial system

The causal analysis of nonconforming product allows to identify each one of the clients in relation to the raw material entering the system (Kg x client) and the operational allocation of machinery in the processing of raw material (Kg x MQ), as well as the cause of rework and material involved. An example of the above can be seen with the client number 2 - dark blue, where the company specifically uses the JCT16 and BR11 machinery for this client to process their orders. Within the exposed analysis, the main causes of the total number of reprocesses for the list of machinery of the organization, exposing the main causes in their order: dye (31.9%), dizziness (19.3%), tone (10.3%), other causes (9.8%), stains (7.2%) and so on in decreasing order. As an added value of the present investigation, it is identified, at the beginning of the main production process, which are the raw materials that are affected because of the associated reprocesses and the causes, generating the following results: Polycotton (31.8%), polyester (15.9%), cotton (15.2%), polycotton-lycra (14%), among others. This type of industrial behavior in the prescriptive analysis of the causes of nonconformities allows the organization to take two types of actions: a) corrective actions, in the identification of the causes of the reprocessed kilograms, generate an action plan to mitigate the impact generated, as well as control and monitoring in the short and medium term; b) preventive actions, in the identification of affected raw material and generation of action plans for production schedules and levels of affectation in order orders.

4.3.3. Analysis of Losses Due to Operational and Non-Operational Causality

Within the industrial behavior of the company, the need to monitor, control and improve the main direct operational causes of the business (operational causality) was identified, relating the main production factors that impact performance and total productivity. All the above was represented through Power BI software under the option – bubble diagram, which illustrates the magnitude in quantity of the causes of equipment non-compliance. All the above is presented in Figure 11:

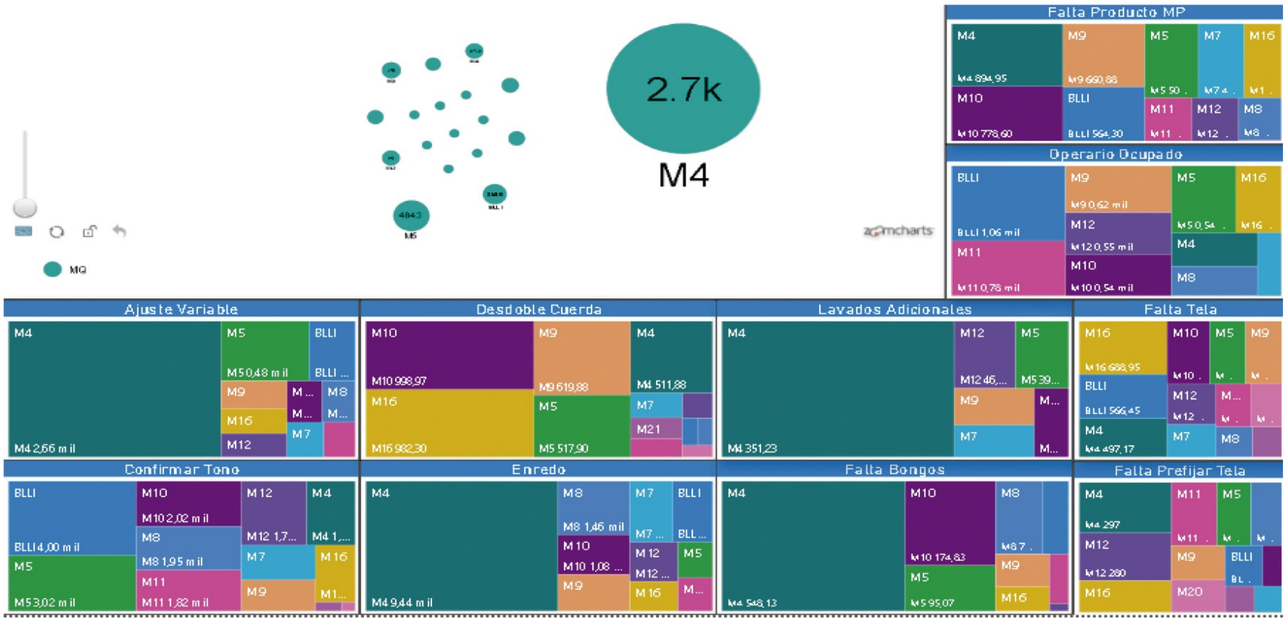

Figure 11. Operational and non-operational causality

In relation to the previous diagram, the machines present in the system (M1-M20) and the number of kg reprocessed in each of them can be evidenced, according to the upper left diagram. According to these results obtained, the machines M4, M5 and M1 are those that represent the greatest number of reprocesses in the system for the period analyzed with resulting figures of 2.7 thousand, 0.484 thousand and 0381 thousand respectively. According to the previous results, this type of machinery should be reviewed, due to the high number of reprocesses that they represent for the system, in order to propose the actions to be taken to avoid this type of damage; however, if an inferential analysis is carried out, such as the causality of machine 4 (M4) for each category of rework, as presented in the block diagram at the bottom of the graph, it can be seen that the main causes of reprocesses for this type of machines are variable adjustment (2.66 thousand), lack of raw material (894.95 thousand) and unfolding of rope (511.78 thousand) and among others, which allows identifying that within in this type of process, there are faults associated with the internal programming of the machine, the absence of material and supplies in the production process, configuration problems and shortcomings in the orders, respectively. This type of identification in the operational analysis must be carried out for each industrial equipment, in search of the generation of immediate action plans and root-cause analysis that allow minimizing the losses present in the system.

A second analysis is performed for indirect operational causality, as can be seen in Figure 12, involving the departments of: quality, logistics and transportation, accounting, and management. Within the results obtained, it is identified that the largest number of reprocesses of non-operative cause are found in M5, M8 and M9 with results of 2 thousand, 1.5 thousand and 1.2 thousand respectively. Considering the previous results, the process of inference of causality is carried out and it is found that the causes of rework, such as in the case of machine 5 (M5), the main one involved in the losses of this category, are in damage of machine, lack of water vapor energy and problems of the Work Orders (OW) and in relation to the present results it is essential that the organization carry out an action plan on the equipment involved, as well as the planning of preventive maintenance and corrective measures of the non-operational causality system. Within the suggested corrections, it is important to consider budget planning and monitoring and control in the short and medium term, in search of better control, support and monitoring of this type of industrial equipment in managerial decision-making.

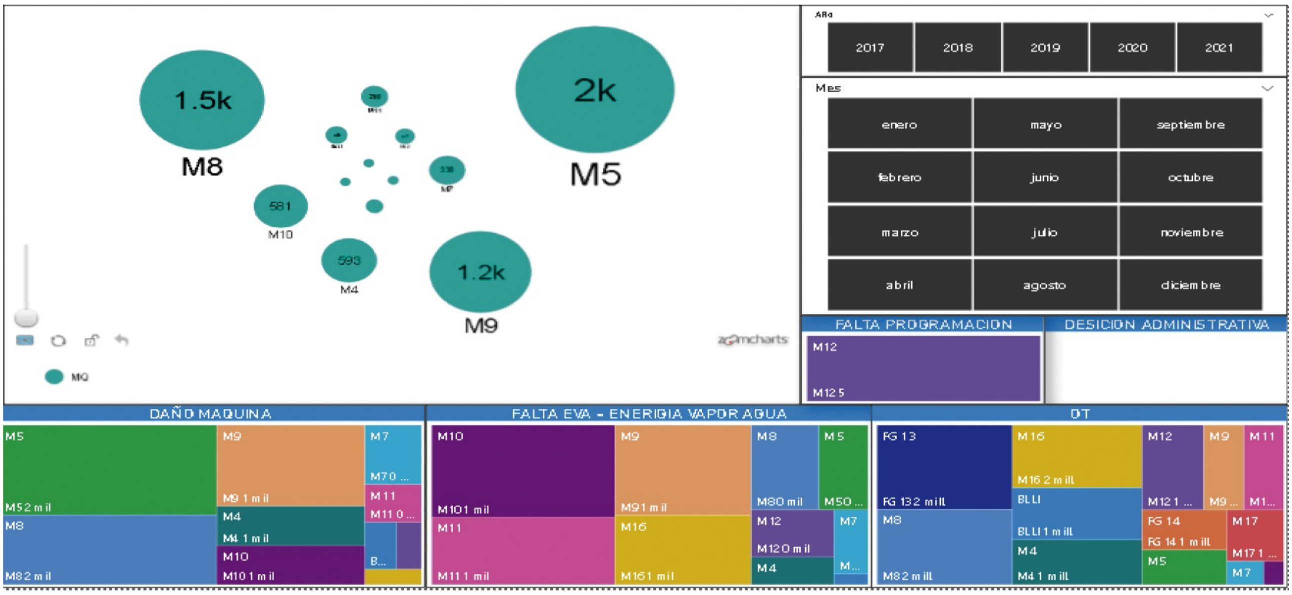

Figure 12. Indirect Operational and non-operational causality

4.3.4. Production Traceability Analysis by Machine

One of the future proposals to be implemented within the present investigation seeks to carry out a predictive analysis and real-time monitoring in relation to the present behavior of the system. Figure 13 illustrates the traceability over the time horizon (2017-2021) of the variables involved in the operating processes of machine 16 (M16). It is possible to identify, within this graph, the machine involved (MQ) in the system, the dry-cleaning process code that must be applied for this order (code), the batch number of the order (Lots-ungrouped), the processing time of each batch (Batch time), the loading efficiency (EFC Load) in percentage and quantity in tons (Efficiency in Tons). If it is analyzed from the timeline cause and effect, a complete monitoring can be presented for each order of each customer order. In the example of Figure 12, it can be seen how the selection of machine 16 (M16) is identified within the stock and codes of the machines in the entire system (MQ count), for the process code of dry cleaner number 31300000, batch number 81628 with a batch processing time of 146 minutes for the related batch, generating a loading efficiency of 96.67%, represented in 109.9 tons.

The type of analysis presented above is an example in the application and implementation of real production systematization and modeling of the industrial study system, which allows the company to review its production batches in real time and generate comparative analysis. of key performance indicators, such as load processing efficiency, batch time efficiency, productivity per batch unit, among others. All the above generates as an output result an improvement in organizational decision making, argued in the systematization of its production systems and support areas that allow a cycle of continuous improvement in production, financial and logistics processes, using tools in the Big Data application as a measure of organizational technological development.

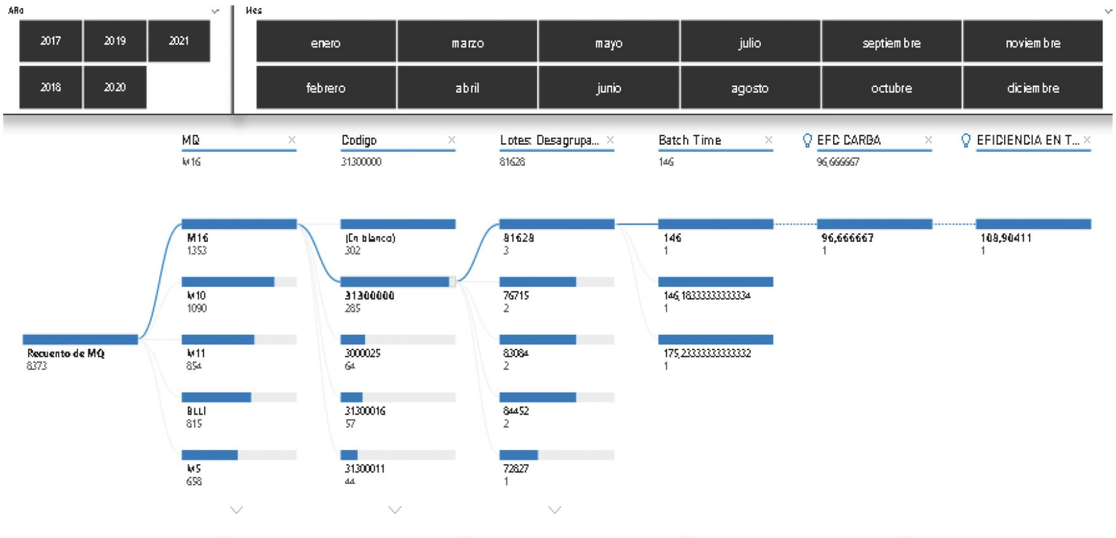

Figure 13. Traceability of production by machine

One more result of this investigation implies to confirm what is described in the literature review facing the high direction importance to assume costs that Big Data implementation comes with, for this study case management and financial production, merchandising and accounting areas executive staff were involved to supply entry and proactive variable's context and relations to analyze output variables models, just as it concludes (Zhou, Liu & Zhou, 2016).

5. Conclusion

With the implementation and use within the proposal of the databases of the fox pro visual system of the production department and the implementation of a model of production and industrial systematization focused on Big Data and applied through Power BI software and visualization of Dashboard generated, the time in organizational decision making for the processing of batches of orders from the company's customers is reduced. The resulting information allows knowing in real time the number of input batches for each customer order, as well as the date of entry and programming in real time in the existing machines, depending on availability; number of batches processed, in relation to the allocation of production programming and projected lead time based on operational requirements; number of output batches, with the assignment of the delivery code, the development of a real-time order tracking system within the warehouses and status is achieved; number of batches reprocessed and the assignment of causes of non-compliance, as well as related equipment and affected materials, among other analysis characteristics, all of the above allows obtaining a real-time improvement of the production line, as well as an increase in profitability and company productivity.

The use of Power BI software with the representation of information in Dashboard in real time allows operators to make decisions in shorter periods of time in an effective and efficient way, argued and justified mathematically in data processed and delivered by the databases of the system. The delivery of information in real time allows a more specific monitoring and control of the system, in relation to the analysis of behavior in real time (descriptive), analysis of future events and operational behaviors (predictive) and analysis of causes for the variables of the system. (prescriptive), improving the traceability of production and associated customer service.

The Big Data implementation experience in the proposed model in this investigation shows, just as it's defined in the literature review, that comes to an evolving technology and is required to continue an innovation process that allows to update, measure and improve even more the proposed model.

The performance of a production and industrial systematization model focused on Big Data and applied through Power BI software and visualization of Dashboard generated, the time in organizational decision-making for the processing of batches of orders from the company's customers is reduced. The resulting information allows knowing in real-time the number of input batches for each customer order, as well as the date of entry and programming in real-time in the existing machines, depending on availability; several output batches, with the assignment of the delivery code, the development of a real-time order tracking system within the warehouses and status is achieved; some sets reprocessed and the assignment of causes of non-compliance, as well as related equipment and affected materials, among other analysis characteristics, all of the above allows obtaining a real-time improvement of the production line, as well as an increase in profitability and company productivity.

Textile production systems generate many output data (structured and semi-structured information) that require proper planning, use, monitoring and traceability in the short, medium and long term. the objective of this research is Develop a sustainable manufacturing proposal for the industrial textile sector with a focus on Big data (entry, transformation, data loading and analysis) in organizational decision making, in search of time and cost optimization and environmental impact mitigation related.

The human-machine interaction of the future seeks to locate the human resource as a decision-making element, where the availability of industrial information in real time is guaranteed.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

References

Abreu, M.F., Alves, A.C., & Moreira, F. (2017). Lean-Green models for eco-efficient and sustainable production. *Energy*, 137, 846-853. https://doi.org/10.1016/j.energy.2017.04.016

Arzamasova, N., Schaler, M., & Bohm, K. (2018). Cleaning antipatterns in an SQL query log. *IEEE Transactions on Knowledge and Data Engineering*, 30(3), 421-434. https://doi.org/10.1109/TKDE.2017.2772252

Aspin, A. (2016). Extending the Data Model with Calculated Columns. *Pro Power BI Desktop*, 205-237. https://doi.org/10.1007/978-1-4842-1805-1_7

Baines, T., Ziaee-Bigdeli, A., Bustinza, O.F., Shi, V.G., Baldwin, J., & Ridgway, K. (2017). Servitization: revisiting the state-of-the-art and research priorities. *International Journal of Operations and Production Management*, 37(2), 256-278. https://doi.org/10.1108/IJOPM-06-2015-0312/FULL/XML

Belman-Lopez, C.E., Jiménez-García, J.A., & Hernández-González, S. (2020). Análisis exhaustivo de los principios de diseño en el contexto de Industria 4.0. *Revista Iberoamericana de Automática e Informática Industrial*, 17(4), 432-447. https://doi.org/10.4995/RIAI.2020.12579

Bhalotia, G., Hulgeri, A., Nakhe, C., Chakrabarti, S., & Sudarshan, S. (2002). Keyword searching and browsing in databases using BANKS. *Proceedings - International Conference on Data Engineering*, 431-440. https://doi.org/10.1109/ICDE.2002.994756

Blunschi, L., Jossen, C., Kossman, D., Mori, M., & Stockinger, K. (2012). SODA: Generating SQL for Business Users. *Proceedings of the VLDB Endowment*, 5(10), 932-943. Available at: https://arxiv.org/abs/1207.0134v1

Doko, F., & Miskovski, I. (2020). *Advanced analytics of Big Data using Power BI: Credit Registry Use Case*. Ss. Cyril and Methodius University in Skopje, Faculty of Computer Science and Engineering, Republic of North Macedonia. Available at: https://repository.ukim.mk:443/handle/20.500.12188/8266

Fatorachian, H., & Kazemi, H. (2018). A critical investigation of Industry 4.0 in manufacturing: theoretical operationalisation framework. *The Management of Operations*, 29(8), 633-644. https://doi.org/10.1080/09537287.2018.1424960

Firouzi, F., Rahmani, A.M., Mankodiya, K., Badaroglu, M., Merrett, G.V., Wong, P. et al. (2018). Internet-of-Things and big data for smarter healthcare: From device to architecture, applications and analytics. In *Future Generation Computer Systems* (78, 583-586). Elsevier B.V. https://doi.org/10.1016/j.future.2017.09.016

Gamero, D. (2021). *SQL And NoSQL Databases for Cyber Physical Production Systems in Internet of Things for Manufacturing (IoTfM)*. Available at: https://smartech.gatech.edu/handle/1853/64678

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137-144. https://doi.org/10.1016/J.IJINFOMGT.2014.10.007

Gawusu, S., Mensah, R.A., & Das, O. (2022). Exploring distributed energy generation for sustainable development: A data mining approach. *Journal of Energy Storage*, 48, 104018. https://doi.org/10.1016/J.EST.2022.104018

Giannakopoulou, S.A., Karpathiotakis, M., Gaidioz, B.C.D., & Ailamaki, A. (2017). CleanM: An Optimizable Query Language for Unified Scale-Out Data Cleaning. *Proceedings of the VLDB Endowment*, 10(11). Available at: https://infoscience.epfl.ch/record/230431 https://doi.org/10.14778/3137628.3137654

Guizzi, G., Falcone, D., & De Felice, F. (2019). An integrated and parametric simulation model to improve production and maintenance processes: Towards a digital factory performance. *Computers and Industrial Engineering*, 137, 106052. https://doi.org/10.1016/j.cie.2019.106052

Guo, Z., Zhang, Y., Zhao, X., & Song, X. (2020). CPS-Based Self-Adaptive Collaborative Control for Smart Production-Logistics Systems. *IEEE Transactions on Cybernetics*, 1-11. https://doi.org/10.1109/tcyb.2020.2964301

Happonen, A., & Ghoreishi, M. (2022). A Mapping Study of the Current Literature on Digitalization and Industry 4.0 Technologies Utilization for Sustainability and Circular Economy in Textile Industries. *Lecture Notes in Networks and Systems*, 217, 697-711. https://doi.org/10.1007/978-981-16-2102-4_63/COVER/

Hernandez, E., Duque, N., & Moreno, J. (2017). *Big Data: una exploración de investigaciones, tecnologías y casos de aplicación*. Available at: http://www.scielo.org.co/scielo.php?pid=S0123-77992017000200002&script=sci_abstract&tlng=en

Jain, A. (2017). Big Data for Supply Chain Management: An Insight to the Analytical Aspects of Tableau & Power BI. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3053915

Khan, M., Wu, X., Xu, X., & Dou, W. (2017, July 28). Big data challenges and opportunities in the hype of Industry 4.0. *IEEE International Conference on Communications*. https://doi.org/10.1109/ICC.2017.7996801

Kleerekoper, A., & Schofield, A. (2018). SQL tester: An online SQL assessment tool and its impact. *Annual Conference on Innovation and Technology in Computer Science Education, ITiCSE*, 87-92. https://doi.org/10.1145/3197091.3197124

Liao, Y., Deschamps, F., Loures, E.F.R., & Ramos, L.F.P. (2017). Past, present and future of Industry 4.0 – a systematic literature review and research agenda proposal. *International Journal of Production Research*, 55(12), 3609-3629. https://doi.org/10.1080/00207543.2017.1308576

Lins, T., & Oliveira, R.A.R. (2020). Cyber-physical production systems retrofitting in context of industry 4.0. *Computers and Industrial Engineering*, 139, 106193. https://doi.org/10.1016/j.cie.2019.106193

López Porrero, B. (2011). *Limpieza de datos: reemplazo de valores ausentes y estandarización*. Available at: http://dspace.uclv.edu.cu:8089/xmlui/handle/123456789/7213

Lu, Q., Li, Z., Zhang, W., & Yang, L.T. (2017). Autonomic deployment decision making for big data analytics applications in the cloud. *Soft Computing*, 21(16), 4501-4512. https://doi.org/10.1007/s00500-015-1945-5

Lu, Y., & Xu, X. (2019). Cloud-based manufacturing equipment and big data analytics to enable on-demand manufacturing services. *Robotics and Computer-Integrated Manufacturing*, 57, 92-102. https://doi.org/10.1016/j.rcim.2018.11.006

Majeed, A., Zhang, Y., Ren, S., Lv, J., Peng, T., Waqar, S. et al. (2021). A big data-driven framework for sustainable and smart additive manufacturing. *Robotics and Computer-Integrated Manufacturing*, 67, 102026. https://doi.org/10.1016/j.rcim.2020.102026

Modoni, G.E., Caldarola, E.G., Sacco, M., & Terkaj, W. (2019). Synchronizing physical and digital factory: Benefits and technical challenges. *Procedia CIRP*, 79, 472-477. https://doi.org/10.1016/j.procir.2019.02.125

Montecinos, E. (2021). *Cuarta revolución industrial y la administración pública en América*. Available at: https://www.redalyc.org/journal/290/29066223002/29066223002.pdf https://doi.org/10.52080/rvgluz93.02

Nadeem, K., Guyer, G.T., Keskinler, B., & Dizge, N. (2019). Investigation of segregated wastewater streams reusability with membrane process for textile industry. *Journal of Cleaner Production*, 228, 1437-1445. https://doi.org/10.1016/j.jclepro.2019.04.205

Nethravathi, B., Amitha, G., Saruka, A., & Bharath, T. (2020). *Structuring Natural Language to Query Language: A Review*. Available at: http://etasr.com/index.php/ETASR/article/view/3873/2407 https://doi.org/10.48084/etasr.3873

Oesterreich, T.D., & Teuteberg, F. (2016). Understanding the implications of digitisation and automation in the context of Industry 4.0: A triangulation approach and elements of a research agenda for the construction industry. *Computers in Industry*, 83, 121-139. https://doi.org/10.1016/J.COMPIND.2016.09.006

Osorio, D., & Vivas, B. (2021). *Estudio de apropiación de las tics e industria 4.0 en los sistemas de producción del sector textil en Colombia, efecto de la pandemia*. Available at: https://repository.udistrital.edu.co/handle/11349/28748

Patriarca, R., Chatzimichailidou, M., Karanikas, N., & Di Gravio, G. (2022). The past and present of System-Theoretic Accident Model And Processes (STAMP) and its associated techniques: A scoping review. *Safety Science*, 146, 105566. https://doi.org/10.1016/J.SSCI.2021.105566

Pereira, A.C., & Romero, F. (2017). A review of the meanings and the implications of the Industry 4.0 concept. *Procedia Manufacturing*, 13, 1206-1214. https://doi.org/10.1016/J.PROMFG.2017.09.032

Petković, D. (2017). SQL/JSON Standard: Properties and Deficiencies. *Datenbank-Spektrum,* 17(3), 277-287. https://doi.org/10.1007/S13222-017-0267-4

Petrillo, A., Picariello, A., Santini, S., Scarciello, B., & Sperlí, G. (2020). Model-based vehicular prognostics framework using Big Data architecture. *Computers in Industry*, 115, 103177. https://doi.org/10.1016/j.compind.2019.103177

Pinheiro, S.A. (2020). *Potencialidades do Power BI Desktop na análise preditiva*. Available at: https://repositorio.ucp.pt/handle/10400.14/32123

Qi, Q., & Tao, F. (2018). Digital Twin and Big Data Towards Smart Manufacturing and Industry 4.0: 360 Degree Comparison. *IEEE Access*, 6, 3585-3593. https://doi.org/10.1109/ACCESS.2018.2793265

Rojko, A. (2017). Industry 4.0 Concept: Background and Overview. *International Journal of Interactive Mobile Technologies (IJIM)*, 11(5), 77-90. https://doi.org/10.3991/ijim.v11i5.7072

Rousopoulou, V., Vafeiadis, T., Nizamis, A., Iakovidis, I., Samaras, L., Kirtsoglou, A. et al. (2022). Cognitive analytics platform with AI solutions for anomaly detection. *Computers in Industry*, 134, 103555. https://doi.org/10.1016/J.COMPIND.2021.103555

Santoso, L.W., & Yulia (2017). Data Warehouse with Big Data Technology for Higher Education. *Procedia Computer Science*, 124, 93-99. https://doi.org/10.1016/j.procs.2017.12.134

Saura, J.R., Herraez, B.R., & Reyes-Menendez, A. (2019). Comparing a traditional approach for financial brand communication analysis with a big data analytics technique. *IEEE Access*, 7, 37100-37108. https://doi.org/10.1109/ACCESS.2019.2905301

Schoeberl, P., Brik, M., Braun, R., & Fuchs, W. (2005). Treatment and recycling of textile wastewater - Case study and development of a recycling concept. *Desalination*, 171(2), 173-183. https://doi.org/10.1016/j.desal.2004.02.105

Shao, G., Brodsky, A., Shin, S.J., & Kim, D.B. (2014). Decision guidance methodology for sustainable manufacturing using process analytics formalism. *Journal of Intelligent Manufacturing*, 28(2), 455-472. https://doi.org/10.1007/S10845-014-0995-3

Siegel, E. (2013). *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, Or Die.* John Wiley & Sons. https://books.google.es/books?hl=es&lr=&id=dGgs3JKUFQIC&oi=fnd&pg=PA187&dq=Siegel,+E.+(2013).+Predictive+Analytics.+United+States+of+America:+John+Wiley+%26+Sons,+Inc.,+Hoboken,+New+Jersey.&ots=9EYfaZlcEM&sig=cFFR-uvCHwGvc9mnNLaRG508aEw#v=onepage&q=Siegel%2C E. (2013). Predictive Analytics. United States of America%3A John Wiley %26 Sons%2C Inc.%2C Hoboken%2C New Jersey.&f=false

Simos, D.E., Zivanovic, J., & Leithner, M. (2019). Automated combinatorial testing for detecting SQL vulnerabilities in web applications. *Proceedings - 2019 IEEE/ACM 14th International Workshop on Automation of Software Test, AST* (55-61). https://doi.org/10.1109/AST.2019.00014

Singh, R.P., Singh, P.K., Gupta, R., & Singh, R.L. (2019). *Treatment and Recycling of Wastewater from Textile Industry* (225-266). Singapore: Springer. https://doi.org/10.1007/978-981-13-1468-1_8

Sobottka, T., Kamhuber, F., Henjes, J., & Sihn, W. (2017). A case study for simulation and optimization based planning of production and logistics systems. *Proceedings - Winter Simulation Conference*, 3495-3506. https://doi.org/10.1109/WSC.2017.8248064

Sun, Y., Liu, Q., Chen, X., & Du, X. (2020). An Adaptive Authenticated Data Structure with Privacy-Preserving for Big Data Stream in Cloud. *IEEE Transactions on Information Forensics and Security*, 15, 3295-3310. https://doi.org/10.1109/TIFS.2020.2986879

Tsai, W.H. (2018). Green production planning and control for the textile industry by using mathematical programming and industry 4.0 techniques. *Energies*, 11(8), 2072. https://doi.org/10.3390/en11082072

Vaidya, S., Ambad, P., & Bhosle, S. (2018). Industry 4.0 - A Glimpse. *Procedia Manufacturing*, 20, 233-238. https://doi.org/10.1016/j.promfg.2018.02.034

Wan, J., Tang, S., Li, D., Wang, S., Liu, C., Abbas, H. et al. (2017). A Manufacturing Big Data Solution for Active Preventive Maintenance. *IEEE Transactions on Industrial Informatics*, 13(4), 2039-2047. https://doi.org/10.1109/TII.2017.2670505

Wang, J., Yang, Y., Wang, T., Sherratt, S., & Zhang, J. (2020). *Big Data Service Architecture: A Survey*. Available at: https://jit.ndhu.edu.tw/article/view/2261

Wang, Y., Liu, J., He, X., & Wang, B. (2018). Design and realization of rock salt gas storage database management system based on SQL Server. *Petroleum*, 4(4), 466-472. https://doi.org/10.1016/J.PETLM.2017.10.001

Wong, Z.S.Y., Zhou, J., & Zhang, Q. (2019). Artificial Intelligence for infectious disease Big Data Analytics. In *Infection, Disease and Health,* 24(1), 44-48. https://doi.org/10.1016/j.idh.2018.10.002

Zhong, R.Y., Lan, S., Xu, C., Dai, Q., & Huang, G.Q. (2016). Visualization of RFID-enabled shopfloor logistics Big Data in Cloud Manufacturing. *International Journal of Advanced Manufacturing Technology*, 84(1-4), 5-16. https://doi.org/10.1007/s00170-015-7702-1

Zhou, K., Liu, T., & Zhou, L. (2016). Industry 4.0: Towards future industrial opportunities and challenges. *12th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2015* (2147-2152). https://doi.org/10.1109/FSKD.2015.7382284

Zhu, L., Yu, F.R., Wang, Y., Ning, B., & Tang, T. (2019). Big Data Analytics in Intelligent Transportation Systems: A Survey. *IEEE Transactions on Intelligent Transportation Systems,* 20(1), 383-398. https://doi.org/10.1109/TITS.2018.2815678