A Big Data / Analytics Platform for Industry 4.0 Implementation in SMEs

Renan Bonnard, Kléber Magno Maciel Vieira, Márcio da Silva Arantes, Rodolfo Lorbieski, Marcelo Canzian Nunes and André Pierre Mattei

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

May 27, 2019
A BIG DATA / ANALYTICS PLATFORM FOR INDUSTRY 4.0 IMPLEMENTATION IN SMEs

RENAN BONNARD, KLEBER MAGNO MACIEL VIEIRA, MARCIO DA SILVA ARANTES, RODOLFO LORBIESKI, MARCELO CANZIAN NUNES, ANDRE PIERRE MATTEI

SENAI INNOVATION INSTITUTE FOR EMBEDDED SYSTEMS

Abstract - Industrial companies operate in an increasingly competitive international environment. SMEs are among the most weak companies in this context and need to continuously innovate to increase their competitiveness, productivity and the quality of their products. Digital transformation, one of the foundations of Industry 4.0, is therefore fundamental to meeting these innovation challenges. The objective of the study is to present the methodology, development and implementation of a new cloud computing platform to collect, store and process data from industrial SME shopfloors. Companies' manufacturing shopfloors employ more and more connected and intelligent devices producing thousands of data that once computed achieve a high added value. The study presents the architecture for collecting this data and storing it in a Big Data solution, and then processing it with advanced artificial intelligence algorithms and/or optimization techniques. This platform has been developed with the aim of minimizing complexity and costs to facilitate the adoption of the platform by SMEs. The implementation and evaluation of the platform was carried out in three companies from three different sectors of Brazilian industry.

Keywords - Industry 4.0, Smart Manufacturing, Big Data, Cloud Computing, Artificial Intelligence

Résumé - Les entreprises industrielles évoluent dans un environnement international de plus en plus concurrentiel. Les PMEs font partie des entreprises les plus fragiles dans ce contexte et ont besoin d’innover continuellement pour augmenter leur compétitivité, productivité et la qualité de leurs produits. La transformation digitale, l’un des fondements de l’Industrie 4.0 est donc fondamental pour relever ces défis de l’innovation. L’objectif de l’étude est de présenter la méthodologie, le développement et l’implémentation d’une nouvelle plateforme infonuagique permettant de collecter, stocker et traiter les données des ateliers de PMEs industrielles. En effet, les ateliers de fabrication des entreprises ont de plus en plus d’objets connectés et intelligents produisant des milliers de données qui une fois traitées ont une grande valeur ajoutée. L’étude présente l’architecture de prélèvement de ces données et son stockage dans une solution Big Data et ensuite son traitement par des algorithmes avancés d’intelligence artificielle et/ou des techniques d’optimisation. Cette plateforme a été développée en cherchant à minimiser la complexité et les coûts pour faciliter l’adoption par les PMEs. L’implémentation et l’évaluation de la plateforme a été réalisée dans trois entreprises de trois secteurs différents de l’industrie brésilienne.

Mots clés - Industrie 4.0, Fabrication Intelligente, Big Data, Infonuagique, Intelligence Artificielle

1 INTRODUCTION

In many countries Small and Medium-sized Enterprises (SMEs) are the cornerstone of the industrial and manufacturing sector [Schiersch, 2009]. They represent nearly 90% of all companies in Brazil [Bastos da S. Guimarães et al., 2018] and much more in Germany [Bür et al., 2018]. Consequently, they have a key role to play in the adoption of the Fourth Industrial Revolution. Nevertheless, these companies are facing difficulties in implementing Industry 4.0 concepts because of innovation culture and budget issues [Mittal et al., 2018]. On the other hand, the impacts of Industry 4.0 on productivity, cost reduction, control over the production process, customization of production, among others, point to a profound transformation in manufacturing plants [Lasi et al., 2014]. This makes it possible to increase the competitiveness of SMEs and thus to survive and gain new market shares either nationally or internationally [Dalenogare et al., 2018]. According to a survey by the Brazilian Industrial Development Agency (ABDI), the annual estimate of industrial cost reduction in Brazil, based on the migration of the industry to the 4.0 concept, will be at least US$19
billion/year. This saving involves efficiency gains (US$9 billion/year), reduction in machine maintenance costs (US$8 billion/year) and energy consumption (US$2 billion/year).

The Industry 4.0 concept that will be highlighted in the next section covers a large aspect of transformation in the Industry, from the introduction of new manufacturing processes, for instance additive manufacturing [Bonnard et al., 2019], systems integration [Lu, 2017], Internet of Things (IoT) [Boyes et al., 2018], advanced robotics [Toquica et al., 2018] and Big Data analytics [Arantes et al., 2018; Tao et al., 2018]. The work presented in this paper contributes to the development of tools for the collection, storage, and analysis of manufacturing and industrial shop floor data.

In order to convince managers and executives of SMEs to adopt new Industry 4.0 concept some of the requirements are the simplicity of the implementation and adoption as well as a budget compatible with companies capabilities [Mittal et al., 2018]. The objective of this study is to propose a low-cost and easy-to-use platform for data analysis in SMEs. This platform enables valuable information extraction for increasing the quality and productivity of industrial manufacturing processes.

2 Literature Review

2.1 Industry 4.0

This concept was born in Germany in early 2010 and is an initiative of the German government, with the help of industrials and academics (Dalenogare et al. 2018). Other initiatives followed throughout the world, notably in the United States with Industrial Internet [Li et al., 2017], in China with Made in China 2025 (L. Li 2018) or in France with Factory of the Future [Liao et al., 2017].

In order to implement Industry 4.0 concept in industrial companies and SMEs a digital transformation is necessary. For that a study of the German National Academy of Science and Engineering proposed a maturity index and defined the different stages in the Industry 4.0 development path [Schuh et al., 2017]. The Figure 1 highlights the three necessary stages in order to complete the path to the Industry 4.0 implementation.

Before the beginning of the implementation of Industry 4.0 in a SME a requirement is a good process organization with technique like lean manufacturing. The first stage in the Industry 4.0 development path that we defined as an “Informed Industry” aims to collect and highlight data from the company in order to know “what is happening” in production processes. The second stage that we defined as an “Integrated Industry” consists in the process chain optimization and horizontal integration enabling to know “what will happen” in the process chain. Finally, the last stage that we defined as a “Smart Industry” aims to achieve “autonomous processes”.

In order to progress in these stages some key technologies of the Industry 4.0 have to be developed and implemented [Arantes et al., 2018]. According to Lu [2017], for the digital aspect of the Industry 4.0 technologies, mobile computing, cloud computing, Big Data / analytics, and the IoT are the key technologies and are presented in the next subsections.

2.2 Internet of Things

According to Dorsemaine et al. [2015] a definition of IoT is: “Group of infrastructures interconnecting connected objects and allowing their management, data mining and the access to the data they generate.” Connected object also called “smart objects” in the literature [Kortuem et al., 2010] are present at each step of the industrial process chain. In the shop floor, these objects are sensors or actuators that are connected and controlled by programmable logic controller (PLC) and supervised by supervisory control and data acquisition (SCADA) system.

2.3 PLC and SCADA system levels

Interoperability and quality of smart objects collected data are key issues in the development of an efficient digital thread based on data-driven manufacturing [Bonnard et al., 2018; Trappey et al., 2017]. These challenges have to be resolved at PLC and SCADA system levels. PLC is a programmable electronic device designed to automate processes such as machine control within a factory and to control industrial robots, for instance. The PLC receives data from connected objects, which are then processed by a defined program for the control of actuators. SCADA system on its side is in charge of the monitoring and control of the processes. In order to guarantee the interoperability of the PLCs and SCADA networks a connectivity standard is also necessary [Álvares et al., 2018].

2.4 Cloud Computing / Cloud Manufacturing

The National Institute of Standards and Technology (NIST) [Mell and Grance, 2009] defined cloud computing as “a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.” One of the main advantages of cloud computing is the ability to access a wide range of remote services from everywhere (data on demand and anytime) [Lee et al., 2013]. As highlighted by Xu
during the development and implementation of a cloud computing/manufacturing solution some essential requirements has to be satisfied: (i) service-centric, (ii) issues, quality of service, (iii) interoperability, (iv) fault-tolerance, (v) load balancing and (vi) virtualization management. Cloud manufacturing is the manufacturing version of cloud computing. It began to appear in the literature in the early 2010’s [Li et al., 2010] due to the need for complex calculations that cannot be performed on production machines.

2.5 Big Data / Analytics

In 2012, the advisory company Gartner gave a definition of Big data as: “Big Data are high-volume, high-velocity, and/or high-variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization”. With the development of IoT and smart/connected objects more and more data are generated and available at PLC and SCADA system levels. Unfortunately, these systems do not have the capacity to store a large amount of various data, while these data are extremely important for correlations, predictions and optimization of industrial processes. That is why Big Data is a necessary technology for the storage of shop floor information.

Big Data is not an end in itself but allows a later processing of its data called analytics. Analytics brings together a set of mathematical and statistical techniques such as artificial intelligence [Lee et al., 2018], data mining [Agard and Kusiak, 2004] and machine learning [Wu et al., 2017] that will allow the processing of Big Data. Analytics’ objective is to find correlations between process parameters and configurations based on the actual history of what happened in the shop floor contained in the Big Data. Then optimization or artificial intelligence techniques are used to optimize the process planning and/or manufacturing parameters. Finally the results are exhibited on dashboards for human analyze and decision-making [Vilarinho et al., 2018; Yigitbasioglu and Velcu, 2012].

2.6 Existing Industry 4.0 Platform

With more and more research teams working on the themes of enterprise digitalization, several data analysis platform architectures have been proposed in the literature. Tao et al. [2018] proposed a conceptual framework of a Big Data Analytics solution, but the detailed architecture and implementation of the proposal are not presented. Woo et al. [2018] introduces a smart manufacturing platform, which has been only assessed for machining. On the other hand, the platform requires a lot of effort during its implementation due to the lack of genericity of the solution. Ji et al. [2018] proposes a big data analytics based optimisation method for enriched distributed process planning that only focused on machining. Other more specific solutions have also been developed, Lu and Xu [2019] proposes a cloud-based manufacturing equipment that can provide on-demand manufacturing services and Shin et al. [2019] proposes a data analytics platform for energy-efficient process planning. The smart manufacturing solution proposed in this study differs in the genericity of its solution, which does not depend on a context or standard for data exchange and the simplicity of its implementation in SMEs.

3 POSITIONNING OF THE STUDY IN THE MANUFACTURING DIGITAL THREAD

Before starting the development and implementation of a company's data analysis platform, it is important to define how this platform will exchange information with the holistic model of a producing company and its manufacturing digital thread.

In manufacturing companies, there is a pyramidal organization of data management (see Figure 2). Data acquisition is done from the shop floor level to the Enterprise Resource Planning (ERP), and decisions/planning orders come from the top floor (ERP) to the shop floor.

The developed Cloud Computing Big Data Analytics Platform exchanges data with all levels of the enterprise digital thread (see Figure 3). The platform also sends back processed data to the ERP and manufacturing execution system (MES). The framework of the solution and the technologies that have been used for the development and implementation of the platform are presented in the next section.

4 CLOUD MANUFACTURING BIG DATA ANALYTICS FRAMEWORK PROPOSAL

The objective of the proposed Cloud Manufacturing Big Data Analytics platform is to be generic and to be implemented in a variety of contexts and industries. The platform therefore requires a flexible system to collect data from the shop floor from a variety of protocols. Due to the potential multiplicity of
data sources, the volume of information collected, and the complexity of mathematical calculations, the choice was made to have a Big Data and analytics on the cloud. To ensure interoperability and scalability, the architecture was elaborated using the concept of services, where the communication between the components is performed via Representational State Transfer (REST) and Application Programming Interface (API). Figure 4 shows the architecture and how the components and database system process information.

The system has a frontend dashboard component that represents a web interface or mobile application through which the user can interact with the environment. Frontend requests data from the backend component, which is connected to a MySQL database.

The backend represents the intermediate component that has the data organized to be presented by the web/mobile interface. It is powered by analytics which uses the REST communication interface to query and enter data that has been processed by the analytics' AI algorithm.

Big Data represents the component that has a large volume of raw data coming from the client's systems. Big Data must provide a scalable and resilient operational database for real-time analysis. These data are received from the external environment, i.e. the environment where it is generated, through a REST API interface, which is also called a collector. Data are collected from the components of the manufacturing digital thread: Connected objects, PLCs, SCADA, MES, and ERP. These raw data are organized by Big Data techniques to facilitate analytics.

Analytics represents the component that has the intelligence and the business rule demanded. The analytics consumes data from the backend and Big Data structure and can be implemented using several techniques of AI, such as machine learning, data mining, genetic algorithm, and neural networks. The information provided by data analysis allows decision-making support from where action planning can be developed. Next, each technology is presented in more detail and its function is explained.

**Figure 4. Platform for Big Data Analytics**

### 4.1 Shop floor level

At this level are considered all objects of the shopfloor, for example robots, sensors, actuators or tablets. These objects are then integrated into the company's network via Ethernet or wireless solutions. Connected objects and end devices are in principle monitored and controlled by PLCs and SCADA. For devices that are not, their information is sent directly via API to the Big Data platform.

### 4.2 PLCs and SCADA level

Taking into account the importance of data storage in an industrial environment, it is necessary to implement software to tackle with gathering and storing large amounts of information for eventual analysis (Big Data). Figure 5 presents the interface between the PLCs, local storage, and the cloud server. This approach ensures robustness in the case of loss of internet connection, allowing the sensor data not to be lost in this case. Figure 5 below exemplifies how this implementation of supervision and data storage is parallel. It shows the data path to its storage and interaction in the Big Data and analytics applications.

**Figure 5. Scada system architecture**

PLC's are connected in a unique network for communication with the SCADA system. This network for communication between CLPs and SCADA system must ensure interoperability. The platform supports several protocols that
meet this requirement, for instance OPC-UA, Modbus or Profibus. Lines, in the figure, represent the connections via cable, wireless communication or Ethernet implemented in the company. There is a local storage database, in the same computer in order to have a buffer in case of Internet connection loss. Data are then sent via API REST to the cloud server to be stored and analyzed.

4.3 Big Data / Analytics level

In order to be able to store and process the various data coming from the manufacturing digital thread of the company a Big Data solution was implemented (see Figure 6). In order to collect the data from the various source of the company (shop floor, PLCs, SCADA, MES and ERP) a generic and scalable collector has been developed. This collector requires a customization during the implementation of the platform. This collector consists of an API that sends REST call. Aiming to collect data without the interference of the collector solution to the factory production, an application was created that runs in the background as a service for the collection of data from the local database in the factory. For this, the Java 8 technologies were used for the communication between the collector and Big Data solutions and for the solution to run in the background, an executable was implemented using Apache Daemons. Java 8 and Spark Java technologies were used for communication between the Big Data and analytics solutions. Spark Java was chosen for this first version of the platform because the pilot companies were not generating too much data volume. For the implementation of this platform in a company generating much more data, the platform can evolve towards the use of Hadoop and Spark Streaming. Cassandra database was used for data storage. Cassandra was chosen because it is a highly scalable distributed database. As it runs on several nodes, the data is distributed through clusters, where each node contains different data. Since there is no master node, all nodes can respond to any request and there is no single point of failure. This means that if any nodes fall, the data from this node is available on other machines in this cluster, for example.

Big Data works through a REST call, which is sent by the collector, and receives a JSON with a list of objects (results of sensors among other things). Its function is to enter the data from the local database sent by the collector and popularize the Big Data, so that it can be used later. When examining large volumes of data to discover hidden patterns, unknown correlations and possible predictions, Big Data analytics is frequently used. The potential of analytics is realized when the decision making process is leveraged through its use. More and more companies are looking for efficient ways to transform large and varied volumes of data into powerful insights. According to Gartner consulting firm, this segment can be basically divided into 4 different levels of analysis, namely: descriptive, diagnostic, predictive and prescriptive.

The intelligence or analytics system uses various artificial intelligence techniques combined with mathematical models and statistics to find correlations in the data and make the proposition of optimized parameters. The main modules of analytics are:

- AI algorithms are used to perform the search for the productivity, balancing or quality optimization of the processes.
- Multi-agent systems are used to simulate the behavior of each equipment of the shop floor and thus be able to represent the behavior of the processes taking into account the interaction between these various agents (equipments).
- Mathematical modeling are used to give the theoretical basis of the operation of each equipment (agent), the more precise and realistic is the model, the more reliable is the result.
- Correlations and regressions: the correlations found in the data made it possible to assemble regressions and thus calibrate the mathematical models to better represent the functioning of the real equipment of the shop floor, thus making the mathematics adaptable and better representative of this reality.

4.4 Backend/Frontend levels

From the data collected, stored and processed by Big Data analytics, the Backend/Frontend system allows managing the users using the system and the dashboards where the data and results of the developed solution are displayed. The Backend consumes the information from the analytics system via API and saves this high value-added data in a MySQL database. The backend has been set up in such a way as to be as generic as possible and to be able to adapt to any industrial context. For this reason its structure is hierarchized and object-oriented.

The four main levels are highlighted on the Figure 7: (i) identification of the company, (ii) identification of the production line, (iii) identification of the station of the production line and (iv) identification of the object of the station. On the other hand, the results of analytics are stored.
with a time stamp and can be presented in the form of a gauge or chart. The Frontend consumes the data of the Backend as a service-based application (see Figure 8). On its side, the Backend provides all the key services for running Frontend dashboard: mobile and web applications.

---

**Figure 8. Generic Database of the Backend**

The Frontend is composed of the following components:

- **Web**: the web allows remote access to the results of the solution and the management of users.
- **Mobile**: mobile allows access to some of the solution's results and receives alerts (also called "push").
- **Dashboard** (see Figure 9): the dashboard allows viewing some of the solution's results. The idea of the dashboard is to display the information on screens with transitions similar to slide shows. There will be no navigation or iteration of user controls. The dashboards will display the results of the web system. The solution may eventually present different screens that will alternate between each defined time interval in seconds.

---

**Figure 7. Generic Database of the Backend**
5 TECHNOLOGY IMPLEMENTATION AND INDUSTRIAL ASSESSMENT

5.1 Implementation

To simplify its implementation, the platform has been put in a container. A container is a standard unit of software that packages up code and all its dependencies so the application runs quickly and reliably from one computing environment to another. This allows the platform to be deployed on any commercial cloud solution. The chosen container solution was Docker.

The platform proposed in this study has been implemented in three pilot SMEs.

5.2 First use case: Equipement Breakdown Identification

In order to evaluate the efficiency of equipment or global productivity of a plant, it is of fundamental importance to monitor some key performance indicators (KPI). Two well-known indicators are overall equipment efficiency (OEE) for equipment and overall line efficiency (OLE) for production line. The objective of the first case study (see Figure 10) was to measure these indicators, detect equipment breakdown and correlate these breakdown with the decrease in productivity (influence on KPI).

To solve this problem, the different types of breakdown were first identified, OEE and OLE defined and then an in-depth search algorithm was developed implemented in the analytics. All its construction was done on the JAVA SE 8 platform. This project reach the first step of the Industry 4.0 development path, earlier presented.

Productivity gains from the project are not direct gains, but indirect gains because the results allow you, for example, to make decisions to change the layout of the line, the allocations of operators or to make new investments.

Thanks to the results presented by the algorithms developed during the project, the company has already identified changes with potential for increasing productivity by around 5%.

5.3 Second use case: Optimized Process Planning

In order to maximize the productivity of an industrial company, it is important to optimize the process planning. This operation generally consists in minimizing the idleness of the machines and equipment. The objective of the second case study (see Figure 11) was the implementation of an optimized and dynamic process planning for a company that is a supplier of industrial equipment. This new application will enable to sell this dynamic and optimized process planning solution as a service.

The genetic algorithm was the solution used to minimize the idleness of the equipment. All its construction was done on the JAVA SE 8 platform. This solution is based on the theory of evolution to generate better solutions. For its realization it is necessary, besides input data, configuration of genetic algorithms values, a representation of chromosomes and a heuristic to be implemented.

This project reach the second step of the Industry 4.0 development path, earlier presented.

The gain for the company refers to the reduction of the idle time of the lines. Hundred different real scenarios have been compared with the analytics proposal and the results gave an average gain of 10.69% ± 1.8% (confidence interval).
5.4 Third use case: Process Parameters Optimization

In order to maximize the productivity and yield of a process, the solution is to optimize the manufacturing parameters. This operation generally consists in the choice of optimized parameters for each equipment of each step of the process. The objective of the second case study (see Figure 12) was the implementation of a system for suggestion of optimized balanced parameters for a production line.

The genetic algorithm and a multi agent techniques were solutions used to optimize the parameters of the process. All its construction was done on the JAVA SE 8 platform. This solution is based on the theory of evolution to generate better solutions. For its realization it is necessary, besides input data, configuration of genetic algorithms values, a representation of chromosomes and a heuristic to be implemented.

This project reach the second step of the Industry 4.0 development path, earlier presented. The gain for the company refers to productivity/yield improvement. Nineteen scenarios of the plant were considered in the period from August 1st to September 14th of 2018 and the results gave an average gain of 2.72 % ± 1.26 % (Confidence Interval).

6 CONCLUSION

The study presented a methodology for developing and implementing a cloud computing Big Data Analytics platform architecture for SMEs. This platform, which is generic, simple to implement and low-cost, is very well suited to SMEs that still have difficulty implementing Industry 4.0 concepts. The evaluation and validation of the platform was carried out in three Brazilian companies from different industrial sectors. The first results are very encouraging with a considerable improvement in company performance and a rapid using of the platform by IT and production teams. The implemented projects have concerned the first two Industry 4.0 maturity levels and work is underway to reach level three.

The study also showed that SMEs can quickly and successfully adopt Industry 4.0 platforms composed of complex new technologies. This allows them to remain competitive in a competitive international market and continue to innovate at lower costs.

The platform has not yet been tested at the third step of the Industry 4.0 development path, earlier presented. Works are being done in this direction and one of the limitations identified is the real-time control of systems that will require a data analysis and control system at the edge level instead of cloud level. Other work is also in progress to integrate models for predictive maintenance.

The platform can also be implemented in other non-industrial applications and work is underway on smart city solutions.

7 ACKNOWLEDGMENT

The authors would like to thank all the people from the SENAI Innovation Institute for Embedded System who contributed to the projects as well as the pilot companies for their support during the evaluation and implementation of the platform.

8 REFERENCES


