25th ABCM International Congress of Mechanical Engineering
October 20-25, 2019, Uberlândia, MG, Brazil

COB-2019-1880
BIG DATA ACQUISITION IN INDUSTRY 4.0: A DISCRETE EVENT MODELING

Felipe Alcarpe Coda
Escola Politécnica da Universidade de São Paulo
coda.felipe@usp.br

Diolino J. Santos Filho
Escola Politécnica da Universidade de São Paulo
diolinos@usp.br

Fabrício Junqueira
Escola Politécnica da Universidade de São Paulo
fabri@usp.br

Paulo Eigi Miyagi
Escola Politécnica da Universidade de São Paulo
pemiyagi@usp.br

Abstract. *Industry 4.0 assumes an autonomous and high-level interaction between machines, connecting Cyber-Physical Systems (CPS) through the Internet of Things (IoT). This requires an ability to collect and to analyze a large volume of data (considered "big data") to improve all the aspects of production processes. In this context, it is essential to understand the process of acquiring this mass of data and its implementation in an Industry 4.0 environment. Thus, the present work focuses on the modeling of a data acquisition system (DAQ) where the set of processes is approached as a discrete event system. The Production Flow Schema (PFS) is the technique adopted for the conceptual modeling of DAQ, since it allows deriving a functional specification in Petri net for the analysis and functional specification of the data acquisition system.*

Keywords: *big data, cyber-physical system, data acquisition, production flow schema, Industry 4.0.*

1. INTRODUCTION

The fourth industrial revolution (i.e. Industry 4.0) is the transformation of traditional plants into smart factories (Kang et al., 2016). In addition to the manufacturing and process industries, these factories involve associated services ranging from logistics to the design and production of goods up to their disposal. This revolution is the product of the Internet of Things (IoT) and Cyber-Physical Systems (CPSs) that combine Information and Communication Technologies (ICTs) for the integration of physical and digital worlds (Pisching et al., 2018) (Langmann and Rojas-Penã, 2016).

IoT is an infrastructure where physical and virtual "things" have identifiers and are integrated composing communication networks (Kranenburg, Dodson, 2008). Likewise, Lee J. et al. (2015) define IoT as a network capable of joining, separating, synchronizing and organizing information from different sources within a factory.

A CPS is a system of a collaboration of computational entities that are intrinsically connected with the physical world and its processes in progress, concurrently providing and using, data access and data processing services (Monostori, 2014). Furthermore, Liu and Jiang (2016) describe how the progress of ICTs has boosted the development of advanced sensors, wireless communication devices, and distributed computing solutions over the last decade. The integration among detection, actuation, monitoring and control devices, and communication technologies allow a real-time "virtual image" of the process control devices.

In this setting, it is clear that the volume of data collected and to be processed is of an order called "big data" and data acquisition systems (DAQ) play a fundamental role, therefore, its components and functionalities must be properly specified.

Lei et al., (2018) describe the process of data acquisition as the capture and storage of relevant datasets of the different processes of the productive systems. In addition, the authors add that a DAQ is generally composed of detection devices (e.g. limit switches, accelerometers, acoustic sensors, thermometers, current sensors, strain gages, etc.), data transmission devices and data storage devices.

In summary, a DAQ is a system in which physical signals of the processes of interest are converted into data according to the following steps (M. C. Corporation, 2012):

1. a variable from a physical element produces a physical signal;
2. the physical signal is converted into an electrical signal by a sensor or transducer;
3. the electrical signal that, in practice, always includes some level of noise goes through a device for signal conditioning and noise removal;
4. the resulting signal is transmitted to an analog / digital converter;
5. the digital signal is converted into a digital data to be stored and processed in a computer for analysis and decision making.

Figure 1 shows the diagram for data acquisition and control of systems, illustrating the path traveled by the collected signal of a physical element up to its conversion into data and processing by a control device. The lower part of the figure illustrates the flow of actuation signals from a decision-making of the control realization device.

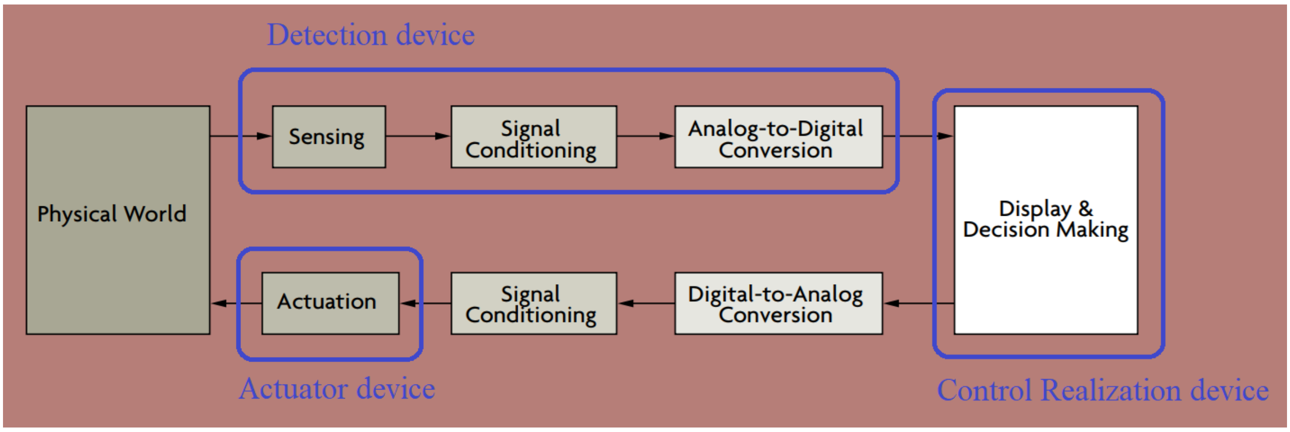

Figure 1. Functional diagram for Data acquisition and control (Adapt from Omega Engineering, 2000).

Over time, detection devices evolved from single-variable electromechanical sensors to electronic systems capable of measuring hundreds of variables simultaneously (M. C. Corporation, 2012). Regarding the concepts of DAQs, the handbooks of Omega Engineering (2000) and M. C. Corporation (2012) detail the fundamentals, conversely, related publications tend to focus on specific applications (Blanco et al., 2014). However, in Industry 4.0 the functionalities involved in the DAQ require a revision of the concepts in function not only of the order of magnitude of the volume of data involved but mainly of what must be extracted and processed to improve the results.

Figure 1 illustrates the basic flow for data collection that, in principle, remains true even in the case of Industry 4.0 and "big data". The analysis of this flow indicates that a DAQ can, therefore, be characterized as a system with a set of attainable states and that the transition between these states is the result of the occurrence of events that can be treated as instantaneous.

In fact, this description is of the operation of a discrete event system (DES) (Miyagi, 1996) (Cury, 2001), and is thus understood that the various methodologies and techniques developed for the study, analysis, and specification of the control of DESs can be exploited to identify the properties of a DAQ, analyze its operational capabilities, and to generate a specification for its implementation. Accurate modeling of DAQ is critical to assess the data collection capabilities described in publications such as (Lee et al., 2015) (Lee et al., 2014) (Sun et al., 2016) within Industry 4.0.

The present work includes 4 sections: Section 1 - an introduction about the context of research and its relevance, Section 2 - concepts associated with this work, Section 3 - modeling of the big data acquisition process, Section 4 - final observations on the study and future work.

2. LITERATURE REVIEW

The main characteristics of the "big data" and how it relates to Industry 4.0 are presented below, followed by the Production Flow Schema (PFS) technique and the methodology adopted for modeling the DAQ.

2.1 Big data

Doug Laney (2001) was the first to define "big data". Its definition unfolded the term in three dimensions: "volume", "variety" and "velocity". Nowadays, authors such as Gil and Song (2016) have added two other dimensions:

"veracity" and "value". There is still a discussion among researchers about the need to add the "value" dimension, but in this work, it is assumed that this represents a characteristic that, in practice, is "very desired" within the data.

A description of each dimension follows:

- “volume”: deals with the amount of data collected (in the order of exabytes, that is, 10^{18} bytes) (Yin, Kaynak, 2015);
- “variety”: deals with the different types of formats and sources of extracted data that can refer to several measurable variables;
- “velocity”: deals with the capability of the data transmission devices to pass the data extracted by the detecting devices to the control devices at a speed compatible with the decision-making needs;
- “veracity”: deals with data reliability and is associated with the availability of tools to ensure this;
- “value”: deals with the information the data contain for an improved perspective (i.e. advantages for the associated processes / services).

According to Lyko, Nitzschke and Ngonga Ngomo (2016), most data acquisition scenarios employ algorithms to guarantee that only the valuable fragments of the data collected are processed. On the other hand, some companies consider most of its data as potentially high-value and sought DAQs capable of handling “big data”. Thus, in Coda et al. (2018), the requirements for "big data" systems, that is, software systems that apply mathematical and statistical techniques to extract data knowledge have been elucidated.

Gittler et al. (2019) contain an approach to data acquisition for analytical applications in Industry 4.0. The concepts addressed by the authors can be compared to our study, pointing out opportunities for improvement. However, publications with more detailed descriptions of the functionalities of DAQs and their implementation in the context of "big data" in Industry 4.0 have not been identified.

2.2 Production flow schema (PFS)

According to Miyagi (1996), Cury (2001), Chung (2004) and Cassandras and Lafortune (2009), a discrete event system (DES) is a way of defining a system with a behavior determined by the occurrence of events that change the state of the system in a discrete way.

Among the various techniques used for the DES modeling, the Petri net (PN) (Miyagi, 1996) (Haustermann, Wagner and Moldt, 2019) stands out in principle for its graphical form of representing processes, and systems and is recognized as useful and effective for a structural and functional analysis. Besides, a PN model is relatively easy to convert into a control program so that it is understood to be an effective technique of functional system specification, that is, it is reliable and efficient as a guide for the practical implementation of the process or the system (object of study).

A PN model is interpreted for each case and different approaches can be adopted which is effective for case-specific analysis and offers greater freedom for the model developer. However, this lack of uniformity compromises the study and development of large and relatively complex systems whereby many actors and different forms and levels of interaction are involved. Thus, in the late 1980s, the PFS (*production flow schema*) was developed and proposed to systematize the development of production systems models (Hasegawa et al., 1988) (Arata and Miyagi, 1997).

The PFS model is an interpreted graph derived from PN to represent processes and systems at different levels of abstraction. As described in Pisching et al. (2018), PFS allows for a gradual detailing of the related functionalities and flows of related items (materials or information) and, because of their intuitive language, generated models can be easily understood by different experts (e.g. engineers, designers or architects).

In the case of DAQ, the PFS is used to identify the data flow in the system, the commands used on the shop floor and the other devices composing the DAQ.

The elements of the PFS are called "activities", "distributors" and "oriented arcs". Activity is a representation of active components, distributor represents passive elements and oriented arcs define the path of the items in the system. Figure 2 illustrates a PFS model with its structural elements.

Figure 2. PFS elements (Pishing et al., 2018).

The type of flow in a PFS model can be classified as one of three types - "primary", "secondary" and, according to Melo et al. (2010) and Junqueira, Villani and Miyagi (2005), a third type can also be considered to represent the "interaction" between components of an information system as request and response operations. Figure 3 shows each flow type.

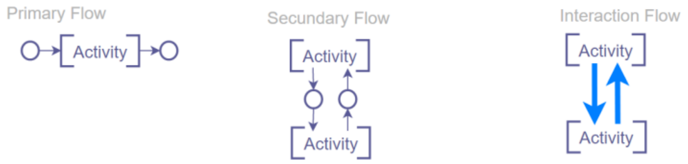

Figure 3. Flow types of a PFS (Pishing et al., 2018).

Figure 4 illustrates the top-down modeling procedure with PFS in which a description at a certain level is gradually detailed until a PN model is obtained.

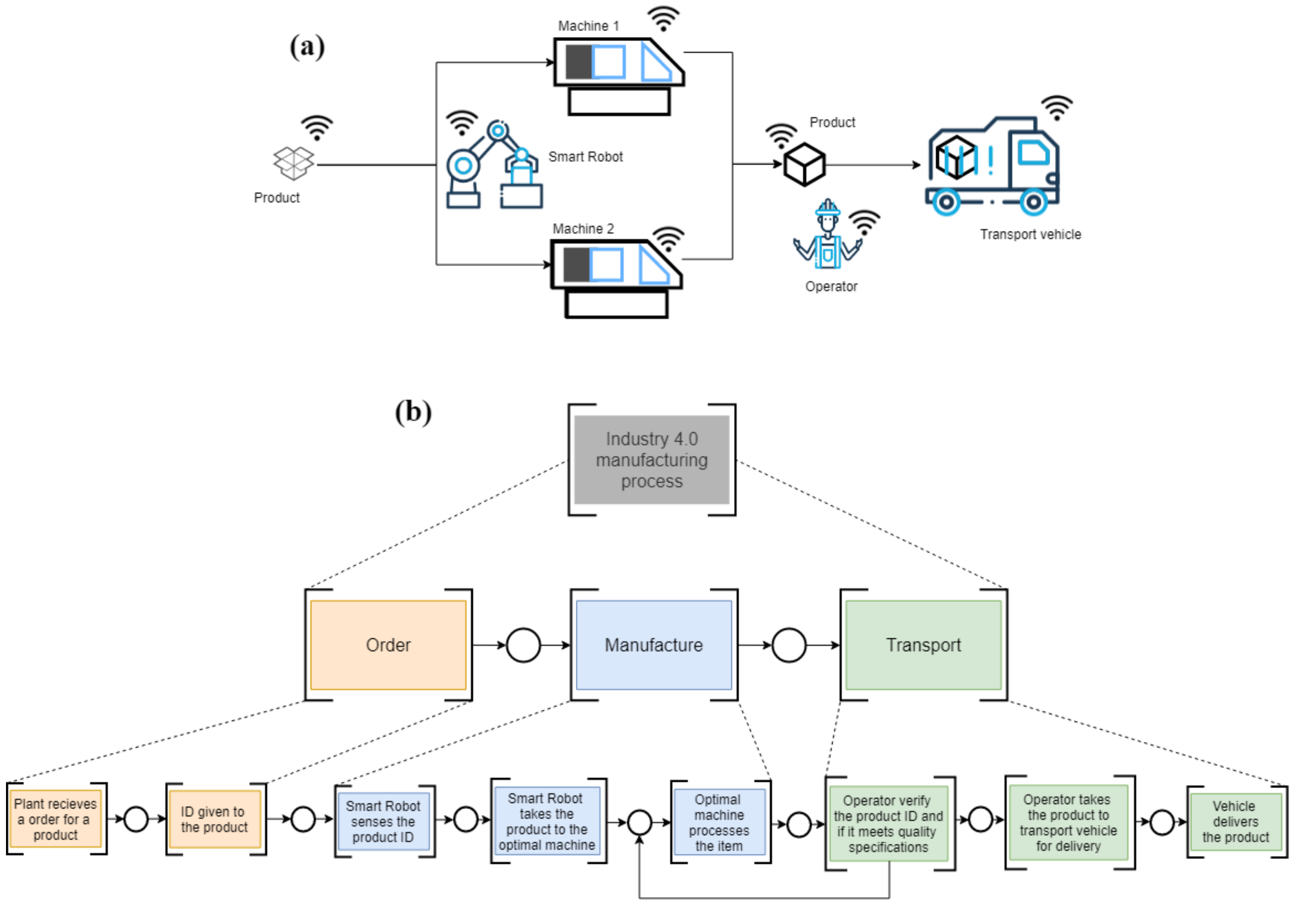

Figure 4. Example of top-down modeling with PFS until obtaining the PN model of the structure and functionality of a manufacturing process in Industry 4.0 where:
(a) is the scheme of the productive system;
(b) are the PFS models of the general structure of the process;

(c)

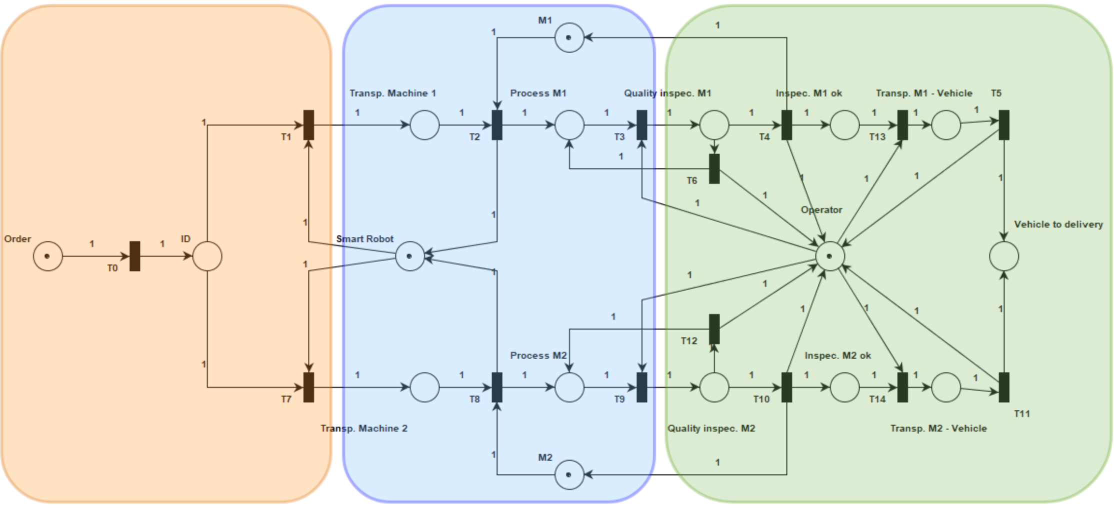

Figure 5. (c) is the PN model of the processes involved.

3. PFS MODELING

To support the modeling of a Big Data-capable DAQ, this work is based on several articles and how they describe the steps in the data acquisition process.

According to Liu and Jiang (2016), the process of industrial data acquisition can have two sources: data collected from **field devices** (i.e. detection devices and machines on the shop floor) in "real time" or "historical system data" collected from **production control and supervision systems** (e.g. ERP, MRP).

The real-time data collected by the detection devices are employed in machine control and process supervision, making these subsystems self-adaptive and self-aware. In addition, the production process as a whole can be optimized by analyzing the data in real time concurrently with historical data from control and supervisory systems. Figure 6 shows the model resulting from this way of interpreting the data usage process.

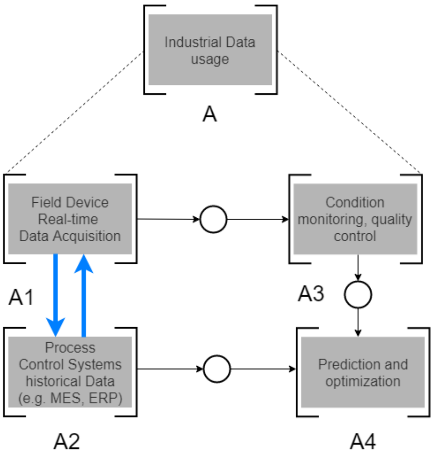

Figure 6. Industrial data usage

Lyko, Nitzschke and Ngonga Ngomo (2016) detail the process of data acquisition in real time. The authors state that data acquisition consists of the process of **collecting**, **filtering**, and **clearing** the data before putting it into a data warehouse or any other **storage solution**.

Figure 7 illustrates the detailing of the real-time data acquisition process.

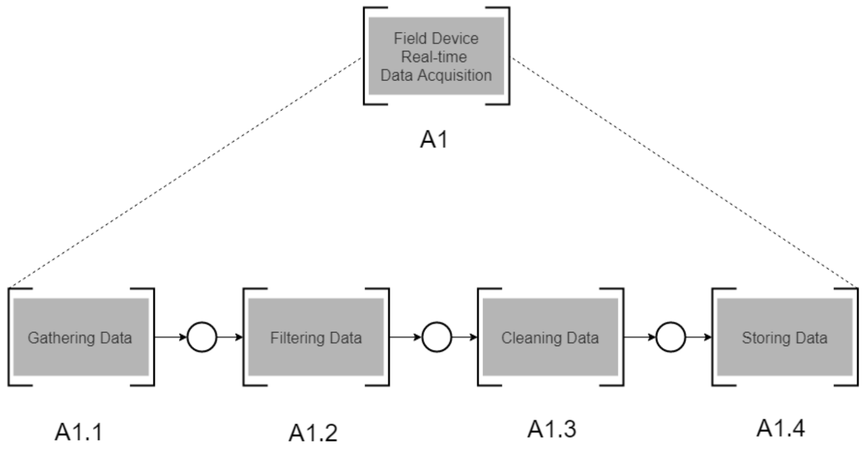

Figure 7. Real-time data acquisition detailing.

According to the manuals, Omega Engineering (2000) and M. C. Corporation (2012), data gathering of field devices consists in using transducers of physical variables. A physical variable is perceived and converted into an analog electrical signal that requires amplification, conditioning, and noise removal. Finally, the signal may be converted from analog to digital (i.e. the continuous signal is measured to a value within a discrete scale). This step allows the signal to be stored and displayed on computers to apply the analysis algorithms.

Figure 8 shows the model generated for the data gathering process.

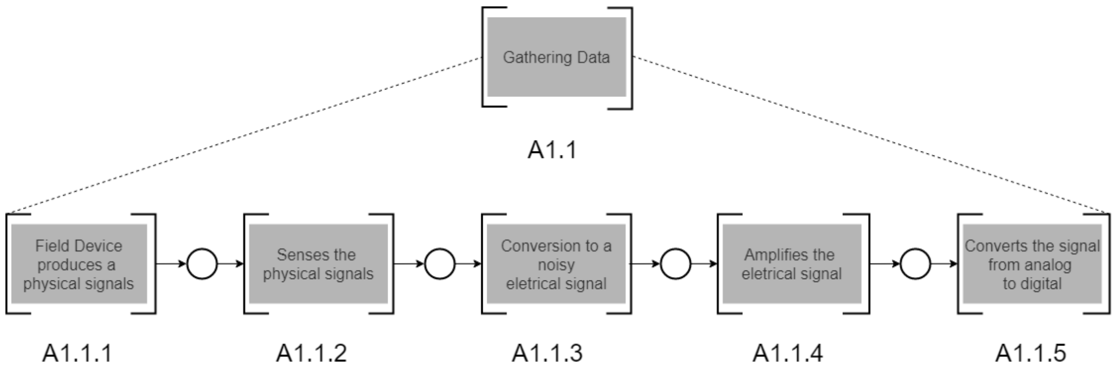

Figure 8. Data gathering detailing.

Nehrey and Hnot (2019) describe the data filtering process as a broad set of strategies for refining datasets. After collecting all the data, at this stage they are integrated and debugged so that only those really necessary are identified, that is, excluding duplicate, redundant and irrelevant data, besides separating those considered confidential, labeling them with rules of access. Figure 9 shows the model of the data filtering process.

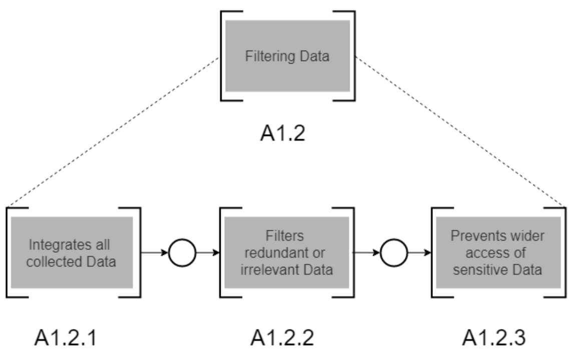

Figure 9. Data filtering detailing.

Cleaning is also a crucial step in preparing the data to be analyzed. A reliable dataset has a higher value than the algorithm employed in its analysis. Practices for data cleansing involve (ELITEDATASCIENCE, 2019):

- Correcting typing or formatting errors (e.g. "N/A" and "not applicable" may indicate different data classes);
- Filter outliers by removing the data considered a suspicious measurement and that do not reflect the actual data;
- Label "gaps" in the data set, referencing missing data.

Figure 10 shows the modeling of the data cleaning process.

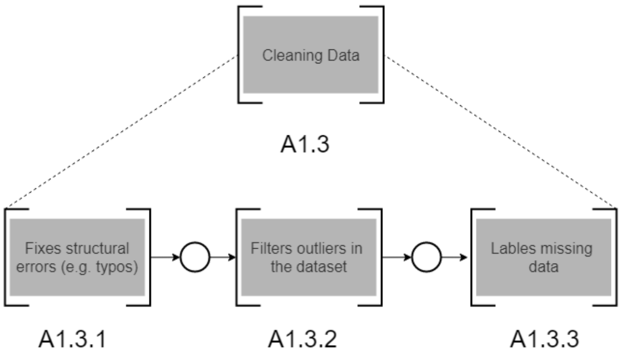

Figure 10. Data cleaning detailing.

A big data "storage solution" concerns the storage and management of data in a scalable form, satisfying the needs of applications that require access to that data. The ideal big data storage system would allow storing an unlimited volume of data, handle high and random-access rates to write and read, flexibly and efficiently handle a wide variety of data models, support both structured and unstructured data and, for privacy reasons, would work with encrypted data (Strohbach et al., 2016).

According to Strohbach et al., (2016), the state-of-the-art in relation to technologies for storing big data is divided into the following types of systems:

- **Distributed File Systems (DFS):** It is a file system such as the Hadoop File System (HDFS), capable of storing large volumes of unstructured data reliably and in relatively low-cost commodity hardware. It is designed for relatively large data files and is suitable for fast data inclusion and mass processing.
- **NoSQL Databases:** The most important systems family of big data storage technologies is probably the NoSQL database management systems. These databases use non-relational[1] data models and do not adhere to atomicity, consistency, isolation, and durability (ACID) properties, which are responsible for ensuring the data accuracy, completeness, and integrity during relational database operations (SQL).
- **NewSQL Databases:** It is a relational database format system that targets scalability comparable to NoSQL databases, maintaining ACID properties (i.e. combining NoSQL speed and performance and SQL reliability).
- **Big Data Querying Platforms:** It is a system with resources that allow queries on large databases such as distributed file systems (DFS) or NoSQL databases. The main feature is a high-level interface (i.e. via SQL) as a language, to ensure low query latencies.

Figure 11 illustrates how the data storage process comprises the use of one or more storage solutions according to the needs of the application.

[1] Non-relational data does not use the traditional data model, relating data in rows and columns tables. Instead, non-relational data is optimally stored for the specific requirements of the stored data type (e.g. data stored in pairs, "key" for identification and "value" for the stored data).

Figure 11. Data storing detailing.

Liu and Jiang (2016) state that historical data on relevant equipment or processes are drawn from corporate control and supervision systems (ERP, MRP, etc.). These systems interact with the "storage solutions" used by the company, bringing together various sets of historical data on productive processes that can be analyzed for prediction and optimization applications. Figure 12 shows the interaction between the "storage solution" and the corporate control and supervision systems.

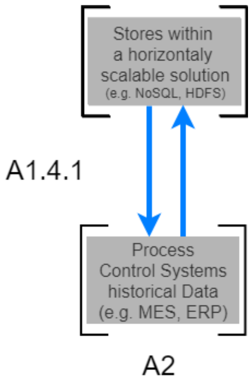

Figure 12. Interaction between corporate control and supervisory systems and the "data storage solution".

Van der Veen, van der Waaij and Meijer (2012) argue that the ideal database format for storing data from detection devices would be a "good" database for single recordings (i.e. whereby each detection device could request one permission at a time to write data to the database), and multiple readings (i.e. each data analysis system would be allowed to perform multiple accesses to read the database).

Figure 13 presents the PFS model of the general big data acquisition process.

Figure 13. PFS model of the general process of industrial data acquisition.

The models obtained were then converted into Petri net models for verification and validation (the presentation of these models is omitted here for space reasons). This study was carried out still at a conceptual level, but this is enough to confirm the structure and the expected functional behavior of the system according to the works cited.

We recognize that not all steps or activities cited here can be followed in practice, but this does not affect the present study whose focus is on the general structure of DAQ processes. In specific applications, there are certain peculiarities which dispense with some activities, that is, the expected functionalities are being carried out in a way other than the usual one with lower cost implementation solutions.

4. FINAL REMARKS

In order to model the Big Data acquisition process, we used the correlated literature that depicts process steps. In interpreting these steps, the goal was to model the process comprehensively, considering the functionalities required for a Big Data DAQ.

The use of the PFS technique for process modeling showed to be effective for visualizing the path of data flows through the system using an intuitive language, facilitating understanding among the different areas of specialists, and systematizing the derivation of the system PN models. Thus, the approach adopted herein was satisfactory for the Industry 4.0 scenario, collaborating to integrate the knowledge of professionals from different areas.

In the context of Industry 4.0, the models obtained are understood as a step for developing systems capable of collecting data that allow reaching the efficiency described by its idealizers. Based on the models obtained, the detailing of the processes should allow an in-depth study of the functionalities and the performance of the processes involved in the acquisition of data so as to have the foundations for the proposal of a DAQ architecture in agreement with RAMI 4.0 (Adolphs et al., 2015) which is the reference model for Industry 4.0 deployment architecture.

5. ACKNOWLEDGEMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal no Nível Superior - Brazil (CAPES), Fundação de Amparo à Pesquisa do Estado de São Paulo - Brazil (FAPESP), Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brazil (CNPq) and Petrobras.

6. REFERENCES

Adolphs, P.; Bedenbender, H.; Dirzus, D.; Ehlich, M.; Epple, U.; Hankel, M.; Heidel, R.; Hoffmeister, M.; Huhle, H.; Kärcher, B.; Koziolek, H.; Pichler, R.; Pollmeier, S.; Schewe, F.; Walter, A.; Waser, B.; Wollschlaeger, M. 2015 "Reference Architecture Model Industrie 4.0 (RAMI 4.0)" *VDI/VDE and ZVEI*. Available at: <https://www.zvei.org/en/subjects/industry-4-0/the-reference-architectural-model-rami-40-and-the-industrie-40-component/>.

Arata, W.M. and Miyagi, P.E., 1997. "Application of PFS Model (Production Flow Schema) Based Analysis of Manufacturing Systems for Performance Assessment". Revista Brasileira de Ciências Mecânicas, Campinas, SP, v. 19, n.1, p. 58-71.

Blanco, J.R., Ferrero, F.J., Valledor, M., Campo, J.C., 2014. "A Low-Cost Open-Source Data Acquisition System" *2014 IEEE 11th International Multi-Conference on Systems, Signals & Devices (SSD14)*.

Cassandras C.G., Lafortune S., 2008 "Introduction to discrete event systems" *Springer*.

Chung, C.A., 2004 "Simulation Modeling Handbook: A Practical Approach". *CRC Press, Taylor and Francis Group*.

Coda, F.A., Salles, R.M., Junqueira, F., Santos Filho, D., Silva, J.R., Miyagi, P.E., 2018. "Big data systems requirements for Industry 4.0". *IEEE International Conference on Industry Applications*.

Cury, J.E.R., 2001 "Teoria de Controle Supervisório de Sistemas Discretos" Apresentado no *V Simpósio Brasileiro de Automação Inteligente, Canela, RS, Brasil*. Available at: <http://www.faatesp.edu.br/publicacoes/controle.pdf>.

EliteDataScience 2019 "Chapter 3: Data Cleaning Steps and Techniques" *Data Science Primer*. Available at: <https://elitedatascience.com/data-cleaning>.

Gil, D., and Song, I., 2016. "Modeling and management of big data: challenges and opportunities" *Future Generation Computer Systems 63*, pp. 96-99.

Gittler, T., Gontarz, A., Weiss, L. and Wegener, K., 2019. "A fundamental approach for data acquisition on machine tools as enabler for analytical Industrie 4.0 applications". *Procedia CIRP*, *79*, pp.586-591.

Hasegawa, K., Miyagi, P., Santos Filho, D., Takahashi, K., Ma, L. and Sugisawa, M., 1999 "On Resource Arc for Petri Net Modelling of Complex Resource Sharing System" *Journal of Intelligent and Robotic Systems, 26(3/4)*. pp.423-437.

Hasegawa, K., Takahashi, K., and Miyagi, P.E., 1988 "Application of the Mark Flow Graph to Represent Discrete Event Production Systems and System Control" *Trans. of the Society of Instrument and Control Engineers Vol. 24 No. 1*.

Haustermann, M., Wagner, T. and Moldt, D., 2019. "Welcome to the Petri Nets World - Frequently Asked Questions with Answers". *Informatik.uni-hamburg.de*. Available at: <http://www.informatik.uni-hamburg.de/TGI/PetriNets/faq/>.

Junqueira, F., Villani, E. and Miyagi, P., 2005. “A Platform for Distributed Modeling and Simulation of Productive Systems based on Petri Nets and Object-Oriented Paradigm” *2005 IEEE Conference on Emerging Technologies and Factory Automation*.

Kang, H.S., Lee, J.Y., Choi, S. Kim, H., Park, J.H., Son, J.Y., Kim, B.H., and Noh, S.D., 2016. “Smart Manufacturing : Past Research , Present Findings , and Future Directions”. Vol. 3, No. 1, pp. 111–128.

Kranenburg, R. van; Dodson, S., 2008 “The internet of things: a critique of ambient technology and the all-see-ing network of RFID”, *Institute of Network Cultures*.

Langmann, R., and Rojas-Peña, L., 2016. “PLCs as Industry 4.0 Components in Laboratory Applications” *iJOE*, vol. 12, no. 7, pp. 37–44.

Lee, J., Ardakani, H. D., Yang, S. and Bagheri, B., 2015. “Industrial big data analytics and cyber-physical systems for future maintenance & service innovation” *Procedia CIRP 38*. pp. 3-7.

Lee, J., Kao, H., and Yang, S., 2014. “Service innovation and smart analytics for Industry 4.0 and big data environment” *Procedia CIRP 16*. pp. 3-8.

Lei, Y., Li, N., Guo, L., Li, N., Yan, T., and Lin, J., 2018. “Machinery health prognostics: A systematic review from data acquisition to RUL prediction” *Mechanical Systems and Signal Processing 104*. Pp. 799–834.

Liu, C. and Jiang, P. 2016. “A Cyber-physical System Architecture in Shop Floor for Intelligent Manufacturing” *9th International Conference on Digital Enterprise Technology - DET 2016 – “Intelligent Manufacturing in the Knowledge Economy Era*. pp. 372–377.

Lyko K., Nitzschke M., Ngonga Ngomo AC., 2016 “Big Data Acquisition”. In: Cavanillas J., Curry E., Wahlster W. (eds) New Horizons for a Data-Driven Economy. *Springer, Cham*. Pp. 39-61.

M. C. Corporation, 2012. “Data acquisition handbook: A reference for DAQ and Analog & Digital Signal Conditioning” *Third Edition* . Available at: < https://www.mccdaq.com/pdfs/anpdf/data-acquisition-handbook.pdf>.

Melo, J., Souit, S., Filho, D., Junqueira, F. and Miyagi, P.E., 2010 “A systematical approach to expose manufacturing system as a service” *2010 9th IEEE/IAS International Conference on Industry Applications - INDUSCON 2010*.

Monostori, L. 2014. “Cyber-physical Production Systems: Roots, Expectations and R&D Challenges” *Procedia CIRP 17*, pp. 9–13.

Nehrey, M. and Hnot, T., 2019 “Cases on Modern Computer Systems in Aviation” *Advances in Computer and Electrical Engineering*, pp.176-190.

Miyagi, P.E., 1996. “Controle Programável - Fundamentos do Controle de Sistemas a Eventos Discretos”. São Paulo, SP: Editora Edgard Blücher, 204p.

Omega Engineering, Inc., 2000. “The Omega Transactions and Master Index Vol. 2”. Putman Publishing Company and OMEGA Press LLC. Available at: <www.omega.com/literature/transactions/Transactions_Vol_II.pdf >.

Pisching, M.A., Pessoa, M.A.O., Junqueira, F., and Miyagi, P.E., 2018. “PFS/PN Technique to Model Industry 4.0 Systems based on RAMI 4.0” In *IEEE 23rd International Conference on Emerging Technologies and Factory Automation*. Torino, Italy, pp. 1153–1156.

Ploner, L., 2014. “Industry 4.0 the next industrial revolution?” *Renesas Edge*. <http://br.renesas.com/edge_ol/global/12/index.jsp, 2014>.

Strohbach, M., Daubert, J., Ravkin, H. and Lischka, M. 2016. “Big Data Storage”. *New Horizons for a Data-Driven Economy*, pp.119-141.

Sun, Y., Song, H., Jara, A.J. and Bie, R., 2016. “Internet of Things and Big Data Analytics for Smart and Connected Communities” *IEEE Access* pp. 766-773.

van der Veen, J., van der Waaij, B. and Meijer, R. 2012 “Sensor Data Storage Performance: SQL or NoSQL, Physical or Virtual” *2012 IEEE Fifth International Conference on Cloud Computing*.

7. RESPONSIBILITY NOTICE

The authors are the only responsible for the printed material included in this paper.