Big Data Acquisition Architecture: An Industry 4.0 Approach

Felipe A. Coda(✉), Diolino J. Santos Filho(✉), Fabrício Junqueira(✉), and Paulo E. Miyagi(✉)

Escola Politécnica da Universidade de São Paulo, Cidade Universitária, Av. Prof. Mello Moraes, São Paulo, SP 2231, Brazil
{coda.felipe,diolinos,fabri,pemiyagi}@usp.br

Abstract. In an Industry 4.0 (I4.0) context there is significant increase in information exchange and storage through the interaction among assets (machines, systems, and people). These data are important because it can lead to the autonomy of assets in decision making. However, the entire organization of I4.0 assets in terms of the quantity and the quality of information to be managed makes the system very complex. Thus, a systematic is needed to deal with this complexity where reference architectures can be used to identify the functionality required to handle this large amount and diversity of data, and how they can be organized. Therefore, the aim here is the specification of a big data acquisition process for its implementation within I4.0 context to ensure quality data for analysis and decision making. The proposed solution is based on reference architectures NBDRA and RAMI 4.0.

Keywords: Industry 4.0 · Big data · Reference architecture · RAMI 4.0 · NBDRA

1 Introduction

Rüßmann *et al.* [1] recognize that I4.0 is a revolution sustained by nine technological pillars. Among them, several researchers such as [2] confirm big data and its analysis as a fundamental aspect.

Regarding the concepts of data acquisition systems (DAQ), manuals like [3] present the main fundamentals. However, in I4.0 the features involved in DAQ require a review, not only on the magnitude of the data involved but mainly of the activities necessary for productive and service processes improvement. An example of said magnitude generated by the connectivity between modern industry, cloud-based solutions and business management is the report [4], which forecasts an increase from 130 to 40,000 exabytes of data generated between 2005 and 2020, and that it doubles every two years onwards [5].

In this context, aiming to implement big data in I4.0, past works [6, 7] centered initially on requirements elicitation. The present work focuses on the development and specification of a system architecture for data acquisition in I4.0, thus, answering the research question of how the big data acquisition relates and is organized within he

Published by Springer Nature Switzerland AG 2020
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2020, IFIP AICT 577, pp. 222–229, 2020.
https://doi.org/10.1007/978-3-030-45124-0_21

I4.0. The architecture proposed here derived from reference architectures for big data systems and the I4.0.

2 Relationship to Technological Innovation for Life Improvement

Brown *et al.* [8] claim that big data will become an asset responsible for optimizing business models, serving as a fundamental basis for competition, and [5] associate big data analysis to enhancements to the development of products and services. These works confirm the expected impact of big data for life improvements that can be extended to I4.0.

Rüßmann *et al.* [1] also evaluates the benefits that I4.0 will bring across four areas:

- **Productivity**: I4.0 is expected to bring improvements in the productivity. The gain in productivity percentage may vary across different industrial areas up to 30%;
- **Revenue growth**: I4.0 is expected to encourage revenue growth. Through the increasing demand for smart equipment and data applications by manufacturers, and the increasing demand for customized products by consumers, pushing an extra revenue growth;
- **Employment**: I4.0 is expected to lead an increase in the employment. As the productive and service systems get more and more complex, I4.0 is expected to further the demand for professionals in the engineering, software development and IT areas;
- **Investment**: I4.0 is expected to increase the amount of investments made in the industrial area. It is expected that in Germany, companies invest €250 billion from 2015 to 2025 to incorporate I4.0 to its production processes.

Within this context, it is recognized the opportunities that big data for I4.0 can bring to productive and service systems, and consequently, carrying technological innovations for life improvements along the value chain[1]. Improvements in analytical capabilities and the quantities and quality of data are also expected to bring about improvements in sustainability (economic, social and environmental) [9].

Consequently, this work has contributions toward said life improvements: fomenting the discussion of big data and specifying an architecture of the big data acquisition activities within the I4.0.

3 Background

In this section, the background related to this work is presented. A thorough review of the big data concepts was discussed in previous work [6].

[1] The value chain is a reference to the set of activities that a company operating in an industrial context performs to deliver a product or service to the consumer market [10].

3.1 System Architecture

An "architecture" is defined as an organizational structure of a system (i.e., an information system plus its devices), containing its parts, relationships, principles and guidelines that serve to its design, implementation and evolution over time [11].

A "reference architecture" is defined by [12] as a document or set of documents that represents the recommended product and service structures and integrations to form a solution (the system architecture), incorporating accepted practices and answering the questions that arise during its development. Adolphs *et al.* [13] argue that reference architectures play an important role in describing key aspects of system structure and are the starting point for developing the tools needed to effectively deploy a system architecture.

3.2 NBDRA

The National Institute of Standards and Technology (NIST) is one of the oldest scientific laboratories in the United States. Among the projects developed at NIST is the NIST Big Data Interoperability Framework, which has generated a 7-volume collection of studies to create a reference architecture that facilitates understanding of the operational complexities of dealing with big data, providing a tool for describing, discussing and developing specific architectures from a common framework [14].

The architecture was developed considering reputable data science companies and ICT (Information and Communication Technologies) solutions.

The NIST Big Data Reference Architecture (NBDRA) considers a big data system with five logical functional components ("**System Orchestrator**", "**Data Provider**", "**Big Data Application Provider**", "**Data Consumer**" and "**Big Data Framework Provider**") connected by interoperability interfaces (i.e., services) and surrounded by two frameworks ("**Security and Privacy**" and "**Management**") that represent the intertwined nature of management and security and privacy in all components.

In this context, the NBDRA provides the structure and necessary functionalities, that is, the background for the big data acquisition enabling its relationship within I4.0.

3.3 RAMI 4.0

The RAMI 4.0 (Reference Architectural Model for Industry 4.0) was developed as reference architecture conceived as a model for systems in I4.0 [15].

Within its three axes, all aspects of I4.0 can be mapped allowing elements to be classified according to this three-dimensional view. The concepts involved in I4.0 can be explored and implemented using RAMI 4.0, enabling a stepwise migration from the present to the I4.0 [15].

According to [16], each axis can be summarized as:

- **Layers:** This axis describe every asset`s technical functions and special properties along its six layers. Through this, all of its characteristics can be virtually mapped;
- **Life Cycle and Value Stream:** This axis characterizes the state of the asset at a specific location at a specific time during its entire life cycle. Through this it is

possible to maintain a record for the life cycle, including the time, location and state parameters, and, at the minimum, the type and instance states;

- **Hierarchical Levels:** This axis is responsible for assigning the asset to the entity responsible for its control. Through this, the intelligence within a machine or system can also be taken in consideration for the decision making along with the control device [13].

A summary of each layer was discussed in previous work [7].

The NBDRA and related literature, is here used to structure the necessary functionalities for the big data acquisition process under the RAMI 4.0.

4 Big Data Acquisition Architecture

In this section, the architecture for big data acquisition with its components and the interactions among them is presented. Additionally, the data flow is detailed, determining the procedure for collecting, integrating, storing and analyzing data. The commands to be sent back to the devices based on the decision making made from the entities that analyzed data are also identified.

4.1 Architecture Components

The proposed architecture elucidates and organize the necessary functionalities in a productive system's context within I4.0. For this purpose, the architecture brings together the functionalities of the components necessary for the acquisition of data organized according to the RAMI 4.0 "layers" axis.

In the first layer, **Asset**, are the physical elements of the process (e.g. equipment, products, sensors, actuators, etc.). This layer references the "Data Provider" discussed in the NBDRA.

In the second layer, **Integration**, are the functionalities responsible for connecting the real and virtual worlds. Also, in this layer is the process of associating the communication technology to the data collected.

In the third layer, **Communication**, data collected from every source are integrated and filtered, removing irrelevant or redundant data. Also, in this layer the communication protocols are associated to the data.

In the fourth layer, **Information**, the data are stored in a solution compatible with its necessities. This layer references the functionalities addressed in the NBDRA as the "Big Data Framework Provider". Also, in this layer are performed the data analysis and visualization of the results, followed by the decision-making process, addressed in as the "Big Data Application Provider" in the NBDRA.

In the fifth layer, **Functional**, are the interfaces for the horizontal integration and the description of all functionalities. This layer is also responsible for generating the rules to be followed and the decision-making logic, addressing the "Data Consumer" in the NBDRA.

In the sixth and last layer, **Business**, are the rules to be followed to ensure that the functions performed maintaining the integrity of every asset. This layer is responsible for the orchestration of the services in the functional layer, addressing concepts from the "System Orchestrator" from the NBDRA.

Figure 1(a) illustrates the architecture components proposed for big data acquisition in the I4.0.

4.2 Interactivity Among Components

Figure 1(b) illustrates the proposed data acquisition process through the interaction of the architecture's components, starting with the occurrence of an event.

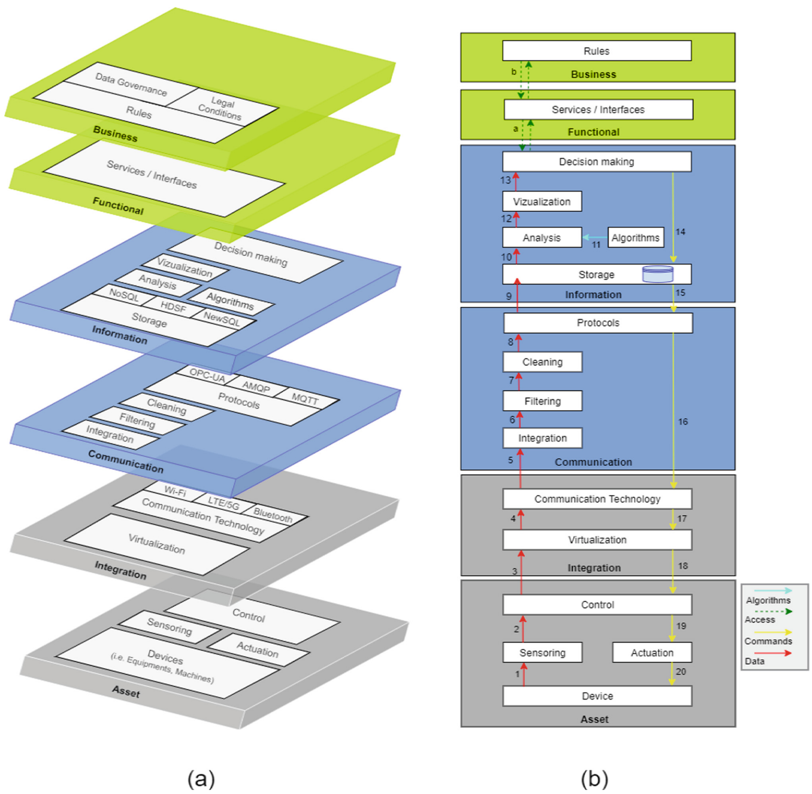

Fig. 1. **(a)** Architecture's components within the "**Layers**" axis of RAMI 4.0 **(b)** Detailing of the data flow among components

The data flow in the architecture establish from the sensoring to the data preparation, the steps for its storage and analysis. The decision-making process can access these data and consult the available services and the rules to be followed. After the decision making, commands are sent back to act on devices.

4.3 Architecture Evaluation

Among the various techniques for modeling systems, the Petri net (PN) 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. In this context, the architecture was evaluated though the use of the PFS/PN (Production flow Schema/Petri net) technique [17].

The PFS model is an interpreted graph that originated from PN to describe processes and systems at different levels of abstraction. PFS allows for a progressive detailing of the activities and flows of relevant items (e.g. materials or data) and, because of their intuitive language, generated models can be effortlessly comprehended by many experts (e.g. engineers, designers or architects). Figure 2 illustrates the PFS model of the data acquisition process along the RAMI 4.0 layers.

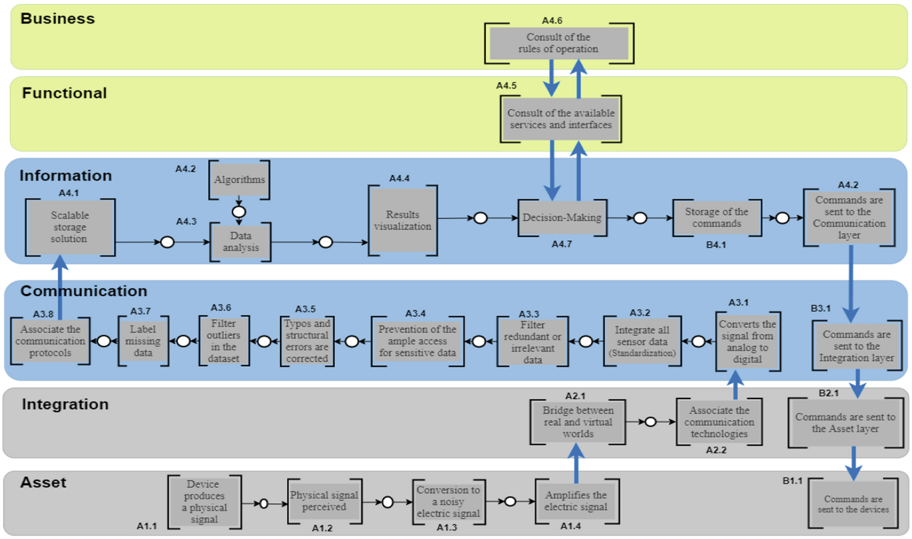

Fig. 2. PFS model of the data acquisition process in the proposed architecture

Through the PFS model it was possible to derive a PN model of the proposed architecture and the interactions among its components. The resulting PN models in turn were formally analyzed and/or simulated, allowing for the evaluation of the architecture structure (e.g. the dynamic of its functionalities such as the existence of deadlocks[2]).

As a sample of the developed models, Fig. 3 illustrates the basic (data) flow and activities (states) of the data acquisition process along the RAMI 4.0 layers. Model

[2] A deadlock occurs when a process goes into permanent standby (i.e. unable to change its state indefinitely) because the resources required by it are being used by another process also permanently waiting [18].

simulations are omitted here for space reasons, since the model has 204 achievable states with 403 situations/scenarios that describe the evolution between those states.

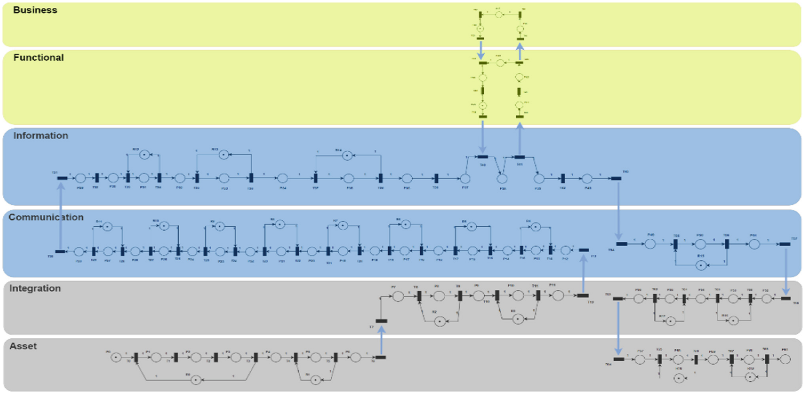

Fig. 3. PN models of the data acquisition process in the proposed architecture

5 Final Remarks

The proposed architecture follows the RAMI 4.0 framework. In addition, the proposed architecture also considers the concepts discussed in other frameworks for dealing with big data, including NBDRA and architectures used by reputable data science companies and ICT (Information and Communication Technologies) solutions. The architecture proposed here is thus based on solutions discussed in both academic and business environments where assured the functionalities associated with the structure and interactivity among the proposed components and the novelty of the topic.

In the context of I4.0, it is noted that many of its aspects are, in fact, still expectations, including RAMI 4.0 which is a work in progress. In this respect, this work is also an application example of RAMI 4.0, addressing the research question of how the big data acquisition relates in I4.0 through its functionalities arranged throughout the RAMI 4.0.

It is noteworthy that, with the proposed architecture, there are possibilities in considering existing data acquisition systems, mapping their functionalities. It is believed that this practice contributes to elucidate existing systems establishing a path for modernization and implementation, as well as its adaptation to the I4.0 paradigms.

Acknowledgement. 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.

References

1. Rüßmann, M., et al.: Industry 4.0: the future of productivity and growth in manufacturing industries. Boston Consult. Group **9**(1), 54–89 (2015)
2. Sun, Y., Song, H., Jara, A.J., Bie, R.: Internet of Things and big data analytics for smart and connected communities. IEEE Access **4**, 766–773 (2016). https://doi.org/10.1109/access.2016.2529723. [s.l.], v. 4, Institute of Electrical and Electronics Engineers (IEEE)
3. Omega Engineering, Inc.: The Omega Transactions and Master Index, vol. 2. Putman Publishing Company and OMEGA Press LLC (2000). www.omega.com/literature/transactions/Transactions_Vol_II.pdf
4. Gantz, J., Reinsel, D.: The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. IDC iView: IDC Analyze Fut. **2**, 1–16 (2012). https://www.speicherguide.de/download/dokus/IDC-Digital-Universe-Studie-iView-11.12.pdf
5. Yin, S., Kaynak, O.: Big data for modern industry: challenges and trends. Proc. IEEE **103**, 143–146 (2015)
6. Coda, F.A., et al.: Big data systems requirements for Industry 4.0. In: IEEE International Conference on Industry Applications (INDUSCON) (2018)
7. Coda, F.A., et al.: Big data on machine to machine integration's requirement analysis within Industry 4.0. In: Doctoral Conference on Computing, Electrical and Industrial Systems, pp. 247–254 (2019)
8. Brown, B., Chui, M., Manyika, J.: Are you ready for the era of 'big data'. McKinsey Q. **4**(1), 24–35 (2011)
9. Watanabe, E.H., da Silva, R.M., Blos, M.F., Junqueira, F., Santos Filho, D.J., Miyagi, P.E.: Framework to evaluate the performance and sustainability of a disperse productive system. J. Braz. Soc. Mech. Sci. Eng. **40**(6), 277 (2018)
10. Porter, M.E.: Competitive Advantage: Creating and Sustaining Superior Performance. Simon and Schuster, New York City (2008)
11. Geraci, A., et al.: IEEE Standard Computer Dictionary: Compilation of IEEE Standard Computer Glossaries. IEEE Press, Piscataway (1991)
12. Hewlett Packard Enterprise: What is a Reference Architecture. Hewlett Packard Enterprise Development LP (2018). https://www.hpe.com/us/en/what-is/reference-architecture.html
13. Adolphs, P., et al.: Reference architecture model Industrie 4.0 (RAMI 4.0). ZVEI and VDI, Status Report (2015)
14. NBD-PWG, Nist Big Data Public Working Group, Reference Architecture Subgroup: NIST Big Data Interoperability Framework: volume 6, reference architecture, version 2. Nist Special Publication 1500-6r1, [s.l.], pp. 1–52 (2017). http://dx.doi.org/10.6028/nist.sp.1500-6r1
15. Hankel, M., Rexroth, B.: The reference architectural model Industrie 4.0 (RAMI 4.0). ZVEI (2015)
16. DIN; DKE; VDE: German Standardization Roadmap Industrie 4.0 - DIN/DKE Roadmap (2016). https://www.din.de/blob/65354/57218767bd6da1927b181b9f2a0d5b39/roadmap-i4-0-e-data.pdf
17. Hasegawa, K., Takahashi, K., Miyagi, P.E.: Application of the mark flow graph to represent discrete event production systems and system control. Trans. Soc. Instrum. Control Eng. **24**(1), 69–75 (1988)
18. Silberschatz, A., Galvin, P.B., Gagne, G.: Operating system principles. John Wiley, Hoboken (2006)