Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 181 (2021) 658–663

Procedia Computer Science

www.elsevier.com/locate/procedia

CENTERIS - International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies 2020

Towards a Data-Centric Architecture in the Automotive Industry

Daniel Alvarez-Coello[a,c,*], Daniel Wilms[b], Adnan Bekan[a], Jorge Marx Gómez[c]

[a]*BMW Research, New Technologies, Innovations; 19 Parkring, Garching 85748, Germany*
[b]*BMW Technology Office Israel Ltd; 121 Menachem Begin Road, Tel Aviv, Israel*
[c]*University of Oldenburg, Deparment of Computing Science; 114-118 Ammerländer Heerstr., Oldenburg 26129, Germany*

Abstract

Vehicle software architectures have been evolving over the last twenty years to support data-driven functionalities. Several enterprises from different domains are currently focusing on improving their data architectures by re-defining the underlying data models to enable core support for analytics and artificial intelligence. Moreover, a common desire to add clear data provenance and explicit context impulses the field of semantics and knowledge graphs. Nevertheless, in the automotive industry, the scenario of connected vehicles implies extra complexity. Vehicle data has an enormous variety, making it essential to develop and adopt standards. This paper presents aspects of ongoing research at the BMW Research Department regarding a conceptual design for vehicle software architectures in the automotive industry. We discuss the principles of a modern data architecture with particular emphasis on the data-centric mindset. We also explore the current challenges and possible working points as the foundation to move from siloed data towards a so-called AI factory.

Keywords: Connected vehicles; data-centric architecture; standardized data; semantic AI; modern data architecture.

* Corresponding author. Tel.: +49-176-2111-7353.
E-mail address: daniel.alvarez@uol.de

10.1016/j.procs.2021.01.215

1. Introduction

Data and analytics capabilities have leaped forward in recent years. The volume of available data is growing exponentially, more sophisticated algorithms have been developed, and computational power and storage have steadily improved [1]. Therefore, we notice that instead of intuition, the new normal is to rely on data to drive digital innovations and business decisions. Indeed, economic analyses consider data as the "most valuable resource" [2]. Nevertheless, we believe that the only way to leverage the full potential is through a modern and scalable data architecture.

This specific data and analytics progress has a significant impact on the automotive industry. Besides all mechanical components, modern connected vehicles incorporate diverse data-driven functionalities. Today's connected vehicles are complex sensor networks on wheels with impressive computation resources and hundreds of sensing devices. In this sense, their software architecture has been evolving with a strong focus on enabling better use of data [3] to support and deliver various applications and services [4]. Consequently, new areas of interest have appeared lately. These areas include application ecosystems, third party integrations, traffic control, safe vehicle road use, driver behavior patterns, vehicle safety features, diagnostics, and others [4,5].

However, such innovation in modern vehicles comes at the cost of enormous complexity [6]. One aspect that stands for this growing complexity is the ever-increasing amount of data and how applications manage it. Since upcoming innovations will depend on how fast the data can be interpreted and consumed and how easy it is to integrate data from different domains, the development and adoption of standards will make the difference [7]. Therefore, to improve and increase data-driven decisions and boost innovation, a suitable data architecture is necessary.

This paper addresses the need for adopting standardized data models towards a data-centric architecture in the automotive industry. In this sense, we present an approach that considers current standard specifications from the World Wide Web Consortium (W3C) and the GENeva In-Vehicle Infotainment (GENIVI) alliance to facilitate web interactions and enable the inclusion of semantics to represent the situational context. This document is organized as follows: in section 2, we discuss the impact of digital transformation on the automotive industry. We present an overview of the principles of a Modern Data Architecture (MDA) in section 3. Then, in section 4, we define the foundation for MDA in association to open standards. Finally, section 5 concludes the article with our final remarks.

2. Digital Transformation

Digital technologies dramatically reshape industry after industry. Many companies are pursuing large-scale change efforts to capture these trends' benefits or to keep up with competitors [8]. Any transformation is challenging, but the digital transformation goes far beyond, requiring changes in both business and organization [9]. It requires a different mindset [10,11] and should be reflected in rethinking the company's overall software architecture.

An example of digital transformation is the one related to the invention of digital photography. It completely transformed the nature and variety of activities connected to photography. Businesses in this domain expanded rapidly, and now we can see photos in many aspects that are part of our daily life (e.g., social media, news, and others). Kodak, one of the most famous companies in the photo industry (which invented the digital camera in 1975), decided to invest further into the traditional business approach and ignored the disruptive change [12]. Thus, no matter how successful a company is, completely ignoring digital transformation can lead to unimaginable losses.

Translating the Kodak example to the automotive industry means that data and analytics are essential, accompanied by exploiting all business potential across domains. We believe that digital transformation should start with a shared understanding of data; this will allow us to redesign existing architectures focusing on four main transformation stages, as reported by Iansiti and Lakhani [9]. These transformation stages provide the guidelines for performing a transition from architectures that have siloed data to ones with a state-of-the-art AI server with a data-centric approach. As shown in Fig. 1, four stages clarify the path from a data and analytics organization to an actual AI factory (i.e., where the potential of data in intense cross-disciplinary activities is seized to a maximum).

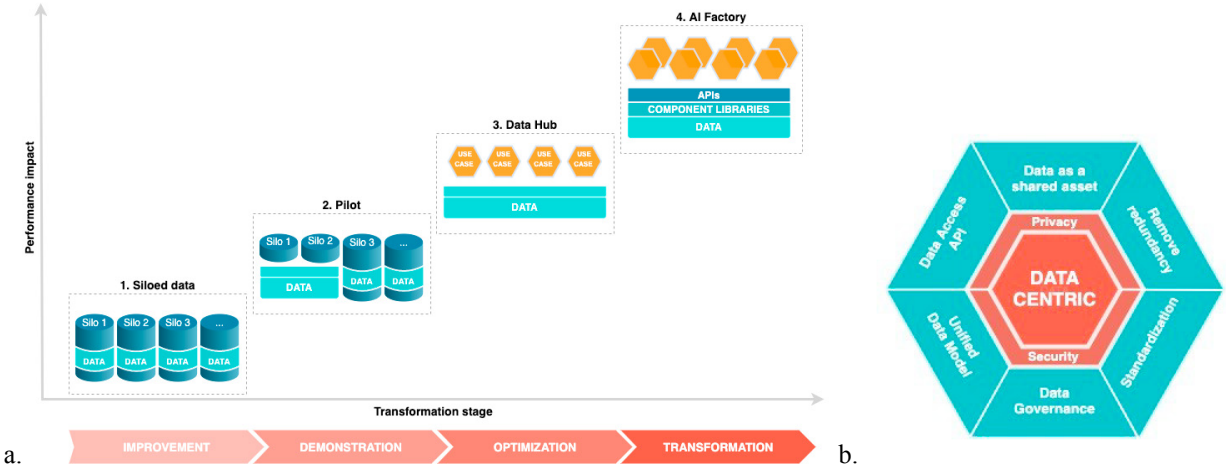

Fig. 1. (a) Stages of transformation to move from an architecture with siloed data towards a state-of-the-art architecture that supports AI and analytics [9]. (b) Some of the aspects involved in a data-centric architecture, where data is the center and most important piece [11].

In stage 1 (i.e., Siloed data), only one group in an organization can access a particular set of data sources. Here, data is mainly use-case centric and not shared across the organization. In stage 2 (i.e., Pilot), a demonstration of analytics' value can be executed without organizational and cultural changes. Consultants and external companies often do this type of demonstration. In stage 3 (i.e., Data Hub), an organization already has to rearchitect and optimize itself to aggregate and label data from many siloed sources. By taking this step, the organization is identifying business opportunities and introducing significant changes. In stage 4 (i.e., AI Factory), the organization develops and adopts a standard operating model for AI, and a major transformation for exploiting the full potential of existing data occurs. Apart from centralized data and architecture that enables reusable software components, this stage also emphasizes the data's clear policies and governance.

3. Modern Data Architecture

A data architecture is composed of "models, policies, rules, or standards that govern which data is collected, and how it is stored, arranged, and put to use in a database system, and/or in an organization" [13]. In order for it to work and be used as intended, we believe that it has to meet the requirements of the data consumers – both within process and usage. Thus, a good data architecture design should be validated with end-consumers' unique requirements, where flows are always defined from the data consumer to the data source.

Additionally, Modern Data Architecture (MDA) addresses the business demands for speed and agility by enabling organizations to quickly find and unify data across its IT infrastructure [14]. In order to achieve the needed flexibility, scalability, and an ideal decoupled infrastructure, one has to focus on several principles of MDA, such as:

- ***Consider data as a shared asset.*** As seen in Fig. 1, data forms the basis for applications. Nevertheless, when data is interpreted, processed, and stored within the application scope only, the same data is handled differently in terms of description, access, and protection. In contrast, thinking of data as a shared asset will enable reusability and is an essential step in the transformation [15].

- ***Look into redundancy and remove copies of the data.*** Data siloes and an application-centric mindset lead to duplication of data, which increases the complexity of the data architecture, especially in terms of data privacy and provenance [10].
- ***Describe data with a flexible unified data model.*** Sharing, interpreting, and processing data throughout the system requires a common understanding. If every application maintains its own data model, the reusability is limited to the vast number of concepts an engineer has to know and understand. Limited reusability increases data redundancy, which leads to more complex data governance. The number of data concepts is directly linked to the complexity of the entire system [11].
- ***Provide the right interfaces for delivering data.*** Well defined and easy to use interfaces are key for reusability and a crucial technical asset of MDA.
- ***Define policies and rules for data governance.*** Policies and rules for data governance have to be in place and enforced. Therefore, at best, the governance process should be directly linked to the technical implementation.
- ***Develop and adopt standards.*** The more open the unified data model is, the higher the chance of reusability in a larger scale. An example of how this could work is the W3C Web Of Things Working Group [16]. Vehicle software architectures are usually brand-specific, and as such, Original Equipment Manufacturers (OEMs) have been developing them to satisfy requirements that are internal to a particular brand or company. In this sense, there are endless combinations of elements, such as transportation buses and protocols. Nevertheless, one aspect that could be improved without affecting brand-specific solutions is the data architecture itself. As shown in Fig. 1, digital transformation puts data as the center and most important piece of the architecture. However, this only works with a common understanding of the data. Thus, the adoption of standards is essential to achieve flexibility, scalability, and a decoupled infrastructure. Furthermore, when many OEMs and the community contribute to the standards, the data correctness will substantially improve thanks to the continuous development to cope with possible imperfections in the data models.

Beyond those fundamental principles, MDA also lives from the customer's trust [17]. Modern and automated processes towards privacy are essential, as is the data security within the entire data chain. Software architectures in the automotive industry continue to evolve, and this transformation process towards a data-centric architecture might be long and challenging. Nevertheless, the benefits of such an architecture will be directly reflected in the overall improvement of the data quality. In this regard, further investigation is needed to find common architectural patterns that define standards that could be adopted by both OEMs and the community.

4. Foundation for a Modern Data Architecture

Typically, enterprises have been using an application-centric paradigm in their software architectures. As McComb points out, allowing applications to modify data models waste resources in the long run [10]. In contrast, the author introduces a new data-centric paradigm [11], which follows particular principles that make data, among other things, self-describing (i.e., with explicit context), and expressed with standards [15]. Looking into the data-centric manifesto [15] and MDA principles, we have found cross needs in data explainability and redundancy. Therefore, we define this as a foundation for a good architecture. One way of handling these challenges can be to employ the well-established hierarchy of Data, Information, Knowledge, Wisdom (DIKW) [18], depicted in Fig. 2.

To discuss the idea in detail, we take a bottom-up approach and start by defining a model for what we consider to be the principal data source in a specific domain. From that point, we can gradually add other components to enhance it. In general terms, the higher the data is represented in this hierarchy, the most explicit context will be directly available for future use. Thus, we aim to climb up the hierarchy with the vehicle data as our starting point and the automotive domain in mind. Further, we will show the importance of contributing to the work of creating well-established standards.

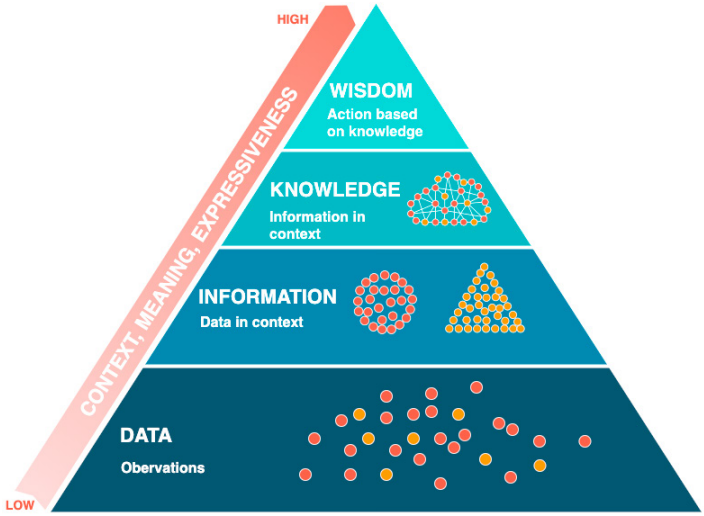

Fig. 2. Data, Information, Knowledge, Wisdom (DIKW) hierarchy [18][19]. Differences between layers are given by the amount of explicit context (i.e., meaning, value, structure, applicability, and expressiveness). Information is defined in terms of domain taxonomies, whereas knowledge covers relationships and interactions between taxonomies from different domains and conceptualize them into a domain ontology.

Towards Information

We start with a domain taxonomy specific to the principal component of the architecture (i.e., vehicle data). Such a taxonomy aims to put isolated vehicle observations from the sensor into context. For that, we make use of the Vehicle Signal Specification (VSS) [20]. VSS is a W3C and GENIVI standard taxonomy for web interactions about vehicle signals. It uses a tree-like structure to explicitly model the membership of attributes, sensors, and actuators to particular branches. This structure is based on the physical structure of the vehicle and defines a common vocabulary for the communication of vehicle data. Further, it adds meta information like the unit, datatype, and description for easy adaptation. Using the same well-described vocabulary as a base of the system directly impacts performance, as discussed earlier. Such a taxonomy can be used directly as the interface towards vehicle data, as shown in other projects [21].

Towards Knowledge

A significant part of the value of using data comes with the integration of heterogeneous sources. Through ontologies, it is possible to define the own domain and link it to others. Therefore, the next step is about using VSS as a base vocabulary for an ontology covering vehicle data and defining the relationships between entities in the automotive domain and others. BMW Research is contributing actively to VSS, and one of the principal contributions has been the extension of VSS into a Vehicle Signal and Attribute ontology (VSSo) [23]. VSSo considers a state-of-the-art light-weight ontology to model semantic sensor networks as Sensors, Observations, Samples, and Actuators (SOSA) [24]. For putting vehicle information into context, the Driving Context Ontology aiming at making sense of cross domain driving event data was also proposed [25]. Additionally, the inclusion of methods to extract the meaning of the sensor observations and their representation in a graph data model is considered in ongoing research [26].

Towards Wisdom

Defining the knowledge layer correctly has a direct positive impact on putting the “knowledge into action” [27]. The wisdom layer relates directly to the data foundation of the fourth step of the digital transformation described earlier. From here, it is possible to formally query the context through the knowledge layer and implement the “action” for a particular use-case.

5. Conclusion

Adopting a modern data architecture is vital for a digital transformation. A fundamental aspect with a significant impact on such transformation is to have a well-defined goal for modeling and handling the data in a system. The number of concepts in a system defines complexity. If this stays unmanaged, it can be seen as a major roadblock on the path towards an AI factory.

We believe that the successful implementation of an automotive taxonomy leads towards this goal. Automotive OEMs rely heavily on the integration of sensors and components produced by Tier-1s. Therefore, we think that a collaborative approach will be the most successful. Undoubtedly, the way towards a successful transformation is long. The only way to succeed is to take one step at a time. The first step is to create the right mindset with a framework at hand to understand the reasoning behind it and guidelines on how to proceed. We believe that the cornerstones of a modern data architecture for automotive presented in this paper can be applied to facilitate the transformation.

References

[1] Henke N, Bughin J, Chui M, Manyika J, Saleh T, Wiseman B, et al. The age of analytics: Competing in a data-driven world. McKinsey 2016;4.
[2] The data economy: Fuel of the future. The Economist 2017.
[3] Haghighatkhah A, Banijamali A, Pakanen O-P, Oivo M, Kuvaja P. Automotive software engineering: A systematic mapping study. Journal of Systems and Software 2017;128:25–55. https://doi.org/10.1016/j.jss.2017.03.005.
[4] Coppola R, Morisio M. Connected Car: Technologies, issues, future trends. ACM Computing Surveys 2016;49:1–36. https://doi.org/10.1145/2971482.
[5] Xu W, Zhou H, Cheng N, Lyu F, Shi W, Chen J, et al. Internet of vehicles in big data era. IEEE/CAA Journal of Automatica Sinica 2018;5:19–35.
[6] Burkacky O, Deichmann J, Doll G, Knochenhauer C. Rethinking car software and electronics architecture. McKinsey & Company; 2018.
[7] Shaheen S, Cohen A, Yelchuru B, Sarkhili S. Mobility on Demand Operational Concept Report. Transportation Sustainability Research Center (TSRC), University of California; 2017.
[8] De la Boutetière H, Montagner A, Reich A. Unlocking success in digital transformations. McKinsey & Company, October, Available Online: [Accessed 10 May 2019] 2018.
[9] Iansiti M, Lakhani KR. Competing in the age of AI: strategy and leadership when algorithms and networks run the world. Boston, MA: Harvard Business Review Press; 2020.
[10] Mccomb D. Software Wasteland. Technics Publications; 2018.
[11] McComb D. Data-centric revolution: restoring sanity in enterprise information systems. 1st ed. Basking Ridge: Technics Publications; 2019.
[12] Scheyder E. Focus on past glory kept Kodak from digital win. Reuters 2012.
[13] BusinessDictionaty.com. Data architecture. Business Dictionary n.d.
[14] Rathi A. Modern Data Architecture. Data Science 2018. https://www.experfy.com/blog/modern-data-architecture (accessed April 29, 2020).
[15] The Data-Centric Manifesto n.d. http://datacentricmanifesto.org (accessed April 29, 2020).
[16] W3C. Web of Things Working Group n.d. https://www.w3.org/WoT/WG/.
[17] Anant V, Donchak L, Kaplan J, Soller H. The consumer-data opportunity and the privacy imperative. McKinsey & Company; 2020.
[18] Rowley J. The wisdom hierarchy: representations of the DIKW hierarchy. Journal of Information Science 2007;33:163–80. https://doi.org/10.1177/0165551506070706.
[19] Sheth A, Henson C, Sahoo SS. Semantic Sensor Web. IEEE Internet Comput 2008;12:78–83. https://doi.org/10.1109/MIC.2008.87.
[20] GENIVI, W3C. Vehicle Signal Specification 2019. https://genivi.github.io/vehicle_signal_specification/ (accessed April 21, 2020).
[21] W3C. VISS implementations 2019. https://www.w3.org/auto/wg/wiki/VISS_implementations (accessed April 29, 2020).
[22] Klotz B, Troncy R, Wilms D, Bonnet C. VSSo - A vehicle signal and attribute ontology. SSN 2018, 9th International Semantic Sensor Networks Workshop, 9 October 2018, Monterey, CA, USA, Monterey, ÉTATS-UNIS: 2018.
[23] Klotz B, Troncy R, Wilms D, Bonnet C. VSSo - A vehicle signal and attribute ontology for the Web of Things. Semantic Web Journal 2019.
[24] Janowicz K, Haller A, Cox SJ, Le Phuoc D, Lefrançois M. SOSA: A lightweight ontology for sensors, observations, samples, and actuators. Journal of Web Semantics 2018.
[25] Klotz B, Troncy R, Wilms D, Bonnet C. A driving context ontology for making sense of cross-domain driving data 2018.
[26] Alvarez-Coello D, Wilms D, Bekan A, Gómez JM. Data architecture for vehicle data semantics and driving context reasoning 2019. https://doi.org/10.13140/RG.2.2.26418.43206.
[27] Ontotext. Semantic Technology Fundamentals 2019. https://www.ontotext.com/knowledge-hub/fundamentals/.