Building an Industry 4.0 Analytics Platform

Practical Challenges, Approaches and Future Research Directions

Christoph Gröger
Robert Bosch GmbH, Robert-Bosch-Platz 1, 70839 Gerlingen-Schillerhöhe, Germany
{firstname.lastname}@de.bosch.com

Abstract The ecosystem of big data technologies and advanced analytics tools has evolved rapidly in the last years offering companies new possibilities for digital transformation and data-driven solutions. Industry 4.0 represents a major application domain for big data and advanced analytics in order to exploit the huge amounts of data generated across the industrial value chain. However, building and establishing an Industry 4.0 analytics platform involves far more than tools and technology. In this paper, we report on our practical experiences when building the Bosch Industry 4.0 Analytics Platform and discuss challenges, approaches and future research directions. The analytics platform is designed for more than 270 factories as part of Bosch's worldwide manufacturing network. We describe use cases and requirements for the analytics platform and present its architecture. On this basis, we discuss practical challenges related to analytical solution development, employee enablement, i.e., citizen data science, as well as analytics governance and present initial solution approaches. Thereby, we highlight future research directions in order to leverage advanced analytics and big data in industrial enterprises.

Keywords data analytics, big data, platform, architecture, citizen data scientist, analytics governance, Industrie 4.0

1. Introduction

The ecosystem of big data and advanced analytics technologies has evolved rapidly in the last years [29]. Machine learning, artificial intelligence and data science offer companies across all sectors new possibilities for digital transformation and data-driven solutions [7]. Industry 4.0 represents a major application domain for advanced analytics and big data in order to exploit the huge amounts of data generated across the industrial value chain and enable, for instance, self-optimizing manufacturing processes and enhanced customer integration [13, 27]. However, building and establishing an Industry 4.0 analytics platform involves far more than tools and technology. In this paper, we report on our practical experiences at Bosch when building an Industry 4.0 analytics platform for a global industrial enterprise and discuss practical challenges, approaches and future research directions.

The Bosch Industry 4.0 Analytics Platform is designed for more than 270 factories as part of Bosch's worldwide manufacturing network enabling, e.g., data-driven process optimization and predictive maintenance. We describe use cases and requirements for the analytics platform and present its architecture. Based on our practical experiences when realizing the platform at Bosch, we investigate major challenges and approaches going beyond purely technological questions. That is, we discuss aspects related to analytical solution development, employee enablement as well as governance and present initial solution approaches. At this, we highlight future research directions in order to leverage advanced analytics and big data in industrial enterprises.

The remainder of this paper is organized as follows: First, we give an overview of Industry 4.0 and data-driven manufacturing as conceptual foundation of our work in Section 2. Next, we present Bosch's business and its worldwide manufacturing network in Section 3 to illustrate the industrial setting. This forms the basis for the design of the Bosch Industry 4.0 Analytics Platform which is detailed in Section 4 regarding use cases, requirements and architecture. In Section 5, we analyze challenges and approaches when realizing the platform and point out future research directions. Finally, we summarize our work and conclude in Section 6.

2. Industry 4.0, Data Analytics and Data-Driven Manufacturing

As conceptual basis of our work, we first give an overview about Industry 4.0 and data analytics in Section 2.1. Then, we detail data-driven manufacturing as a core element of Industry 4.0 in Section 2.2.

2.1. Industry 4.0 and Data Analytics

Industry 4.0, respectively Industrie 4.0, is a general term referring to the next generation of industrial value generation based on the comprehensive use of internet-of-things (IoT) technology and cyber-physical systems [3]. It relates to the fourth industrial revolution after mechanization, electrification and informatization and aims at the complete digital interconnection of all processes and objects across the industrial value chain. The vision is to realize self-optimizing processes and products, to achieve dramatical improvements in productivity as well as agility and to realize novel types of services [5]. Industry 4.0 has its roots in an initiative of the German government and can be seen as a pendant to the US-driven approach of the Industrial Internet [20].

The intense employment of IoT technology and cyber-physical systems across the industrial value chain leads to huge amounts of heterogeneous data comprising, for example, product model data from engineering, machine sensor data from manufacturing as well as telemetry data from product usage [24]. Extracting business insights and knowledge from these data, e.g., for predictive maintenance or manufacturing quality analyses, is one of the major challenges in Industry 4.0 [13, 18]. Thus, Industry 4.0 constitutes a major application domain for data analytics and the goal-oriented use of data analytics techniques represents one of the critical success factors for the realization of Industry 4.0.

In this paper, we use the term *data analytics* as a broad and general term for all data-driven analysis techniques and concepts across enterprise data and big data, because we observe many uncertainties about related terms in our daily work and in current literature [9, 11]. That is, we explicitly subsume techniques and concepts on *business intelligence* [23], *big data* [9], *data mining* [19], *artificial intelligence* [35] and *advanced analytics* [4] under the term data analytics. A *data analytics platform* generally refers to an information system for analytical data processing [23].

2.2. Data-Driven Manufacturing

To structure Industry 4.0 as an application domain of data analytics, we take a product-life-cycle-oriented view on the industrial value chain and differentiate between product development, product manufacturing as well as product usage and recycling [41]. *Data-driven manufacturing* specifically refers to the application of data analytics in product manufacturing [15, 39]. In the following, we focus on data-driven manufacturing as it constitutes the first practical application area of the Bosch Industry 4.0 Analytics Platform.

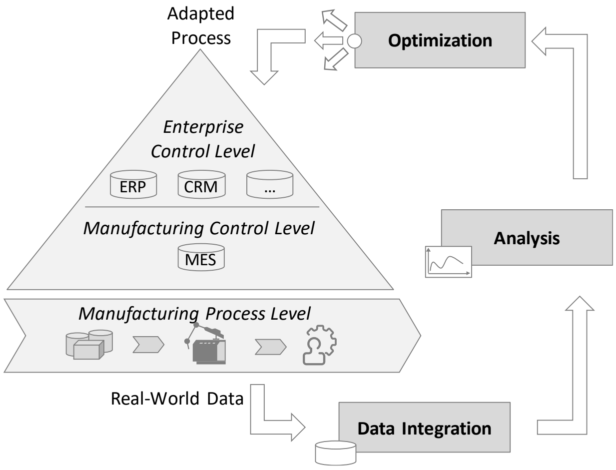

Figure 1: Concept of data-driven manufacturing

On a conceptual level, data-driven manufacturing is comprised of three phases which represent a learning cycle [15]. The goal is to realize continuous data-driven improvement by extracting knowledge and insights from data as follows (see Figure 1):

- The *starting point is an existing manufacturing process and related source data* from various systems across all hierarchy levels of manufacturing, for example, from machines, manufacturing execution systems (MES), enterprise resource planning (ERP) systems or customer relationship management (CRM) systems.
- *Data integration* is about the collection, cleansing, integration and historization of all these heterogenous source data to provide a harmonized data basis for further analyses.
- *Analysis* refers to the comprehensive evaluation of data by applying various data analytics techniques such as metric calculations, reporting or machine learning.
- *Optimization* is about deriving and implementing concrete improvement actions based on the generated analytical results to adapt the process, e.g., by changing specific machine parameters.
- Finally, the *adapted process is executed again* and generates new data as a basis for the next iteration of the cycle.

To leverage data-driven manufacturing, both organizational and IT-technical aspects have to be taken into account, e.g., to replace historically grown time-triggered production planning by an event-triggered concept [39]. In this paper, we focus on IT-technical aspects, in particular, the design of a suitable data analytics platform.

3. Bosch Business and Bosch Manufacturing Network

In the following, we present a short overview of the business of Bosch and its global manufacturing network in order to illustrate the industrial setting of our work.

Bosch is a leading global supplier of technology and services with around 390.000 employees and 73 billion euros sales revenue represented in roughly 150 countries worldwide [37]. The business of Bosch is structured according to four business sectors comprising automotive technology, industrial technology, energy and building technology as well as consumer goods.

The worldwide *manufacturing network of Bosch* spans across all business sectors and is comprised of more than 270 factories. They are located in Europe, the Americas and Asia Pacific and manufacture a broad range of products from sensors for cars and smartphones over battery systems and electrical drives to power tools and solar thermal systems. Thus, there is a huge variety of manufacturing processes from highly automated large-scale production to single-item order production. This leads to an enormous diversity of manufacturing data sources comprising specialized MES and quality management systems, completely different types of manufacturing machines as well as miscellaneous sensor-based measuring systems.

In view of this historically grown and comprehensively optimized manufacturing network, it is one of the strategic objectives of Bosch to apply Industry 4.0 concepts to further on ensure global competitiveness, agility and productivity of its manufacturing capabilities. Enabling data-driven manufacturing is an essential element of Bosch's Industry 4.0 initiative and the corresponding data analytics platform is detailed in the following section.

4. Bosch Industry 4.0 Analytics Platform

The Bosch Industry 4.0 Analytics Platform is a standardized Bosch-internal data analytics platform for Industry 4.0 to exploit the huge amounts of data across Bosch's worldwide manufacturing network and generate valuable business insights. In the following, we first detail use cases and requirements for the platform in Section 4.1 and 4.2. Next, we present the architecture of the platform as well as its deployment in Section 4.3.

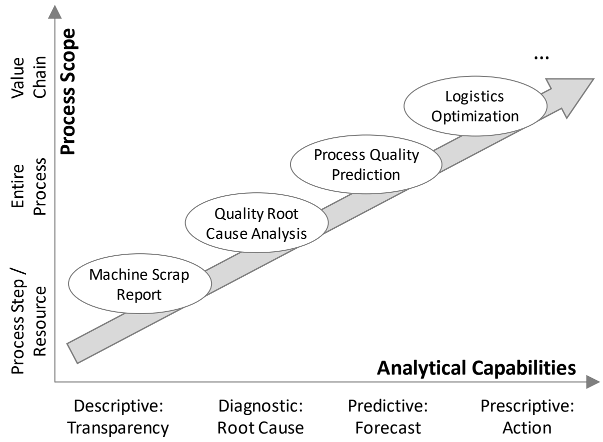

Figure 2: Sample data analytics use cases in data-driven manufacturing

4.1. Data Analytics Use Cases

The Bosch Industry 4.0 Analytics Platform is designed from the beginning on for the entire industrial value chain comprising product development, product manufacturing as well as product usage and recycling. In the following, we focus on product manufacturing, that is, data-driven manufacturing (see Section 2.2), as it constitutes the first application area which we gained comprehensive practical experiences on. The pre- and post-manufacturing phases will be in the focus of subsequent application areas.

There is a huge variety of *data analytics use cases* in data-driven manufacturing which can be structured according to process scope and analytical capabilities (see Figure 2). Regarding *process scope*, use cases can refer to single process steps or resources, to an entire process consisting of several steps or to several processes across the value chain. *Analytical capabilities* are technology-independent and can be differentiated in descriptive, diagnostic, predictive and prescriptive analytics with increasing functional power [21]. *Descriptive analytics* focus on transparency and describe existing data structures typically by calculating metrics, e.g., in a scrap report for a certain machine. *Diagnostic analytics* aim at root cause analyses, for instance, to identify root causes of declining process quality. *Predictive analytics* focus on forecasting, e.g., to predict process quality. *Prescriptive analytics*, the most powerful and most complex form, focus on optimization and generate concrete action recommendations for a defined goal, e.g., to derive optimal logistics procedures minimizing lead times across the value chain.

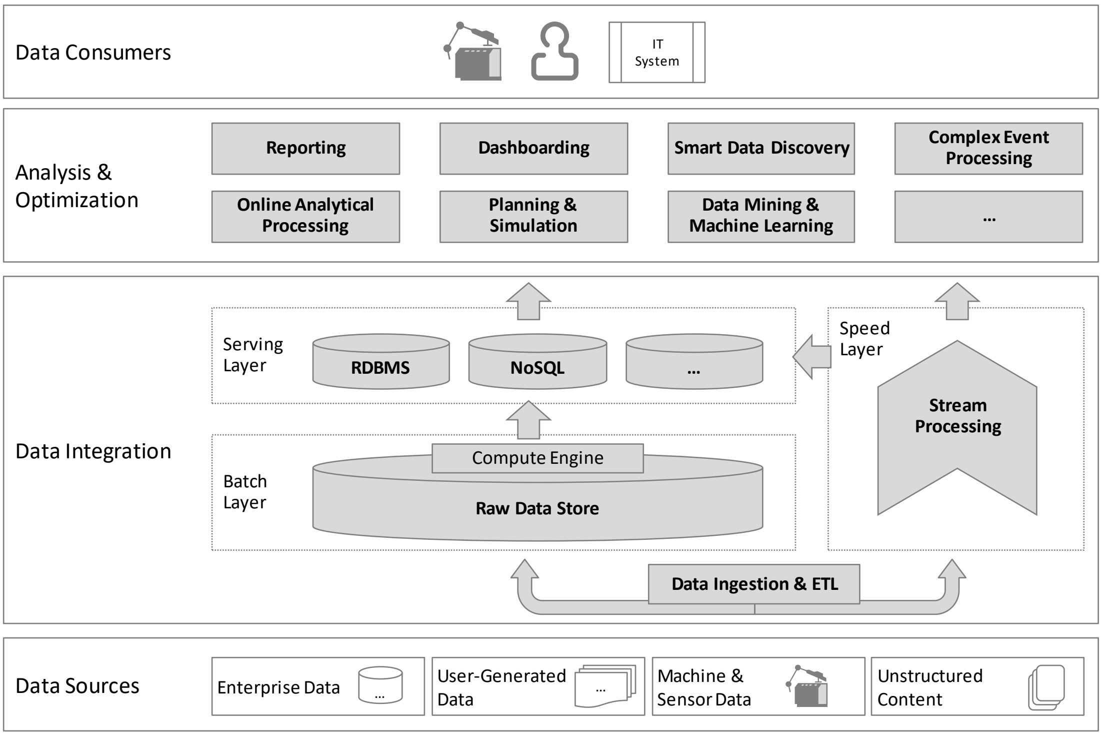

Figure 3: Conceptual architecture of the Bosch Industry 4.0 Analytics Platform

In view of the complexity of Bosch's worldwide manufacturing network, its processes and data sources (see Section 3), we do not rely on single use cases or specific data sources to define the requirements of the Bosch Industry 4.0 Analytics Platform. Instead, we follow a generic approach to design a standardized and generic data analytics platform which is able to address the whole range of data analytics use cases in data-driven manufacturing as described in the following section.

4.2. Requirements

The design of the Bosch Industry 4.0 Analytics Platform is based on *functional and non-functional core requirements* (R_i).

Functional core requirements are defined by data sources, analytical capabilities as well as target groups to be addressed by the platform as follows:

- R_1: The platform should support the whole range of structured and unstructured *data sources*, that is, classical *enterprise data*, e.g., from ERP and MES systems, *user-generated data*, e.g., spread sheets, *machine and sensor data*, e.g., from manufacturing robots, as well as *unstructured content*, e.g., photos taken by manufacturing quality systems [24].
- R_2: The platform should support the whole range of *analytical capabilities*, i.e., *descriptive*, *diagnostic*, *predictive* and *prescriptive analytics* [21].
- R_3: The platform should address the whole range of *target groups* which produce and consume analytical results, i.e., *business users*, *business analysts* and *data scientists* as well as *technical systems* such as machines [28].

Non-functional core requirements refer to essential quality attributes of the platform and are defined as follows:

- R_4: The platform should provide *standardization* of tools and governance concepts to ensure reusability and knowledge transfer.
- R_5: The platform should provide *scalability* across small data and big data to be universally applicable.
- R_6: The platform should provide different *deployment options* to be flexibly implementable.

4.3. Architecture and Deployment

To implement the core requirements, the conceptual architecture of the Bosch Industry 4.0 Analytics Platform is based on the Lamda architecture paradigm [30]. As shown in Figure 3, the architecture comprises all necessary components to realize the three phases of data-driven manufacturing, namely *data integration*, *analysis* and *optimization* (see Section 2.2).

Data from structured and unstructured *data sources* (R_1) is collected and ingested into both the batch layer and the speed layer. The *batch layer* focuses on batch processing of historic data in a *raw data store or data lake*, e.g., for data mining, and aggregates data in the *serving layer*, e.g., to populate a data warehouse with key performance indicators. The *speed layer* facilitates near-real-time data processing based on streaming techniques to circumvent the inherent latency of the batch layer. The speed layer either provides data directly for analysis and optimization on streaming data or feeds data into the serving layer. The latter contains various storage options, e.g., *NoSQL systems* or *relational database management systems* (RDBMS), to provide aggregated or application-specific data. The combination of the three layers facilitates scalability across small data and big data for both batch processing and near-real-time processing (R_5). On top of these layers, various data analytics techniques are applied for analysis and optimization. They range from classical *reporting* and *online analytical processing* (OLAP) over *planning and simulation* to *data mining* and *complex event processing* realizing the whole range of analytical capabilities (R_2) for different types of *data consumers*. Data consumers comprise human user groups, IT systems, e.g., MES, and machines which directly consume analytical results (R_3).

For all of these components, we did a comprehensive tool evaluation to define a standardized and integrated tool stack (R_4). Due to confidentiality reasons, we do not list individual tools in this paper, but we can state that we make use of both open-source tools, particularly from the Hadoop ecosystem [1], and commercial tools.

Regarding the *deployment* of this tool stack, a one-size-fits-all approach with a single platform instance would not fit the heterogenous global manufacturing landscape of Bosch. This is why we provide different deployment options with a centralized operations management (R_6). Bosch-internal customers, particularly individual factories, can order the platform from the central IT division as a dedicated local deployment, make use of a central shared service or request a hybrid setup.

5. Challenges, Approaches and Future Research Directions

Taking into account the generic requirements of the Bosch Industry 4.0 Analytics Platform as well as the heterogenous global manufacturing landscape of Bosch, we were faced with manifold challenges when practically building and establishing the platform. In the following, we explicitly focus on challenges beyond tools and technology to illustrate the wide range of aspects which have to be addressed in order to successfully establish an Industry 4.0 analytics platform.

We present an overview of the challenges in Section 5.1 and further discuss them in Sections 5.2 to 5.4.

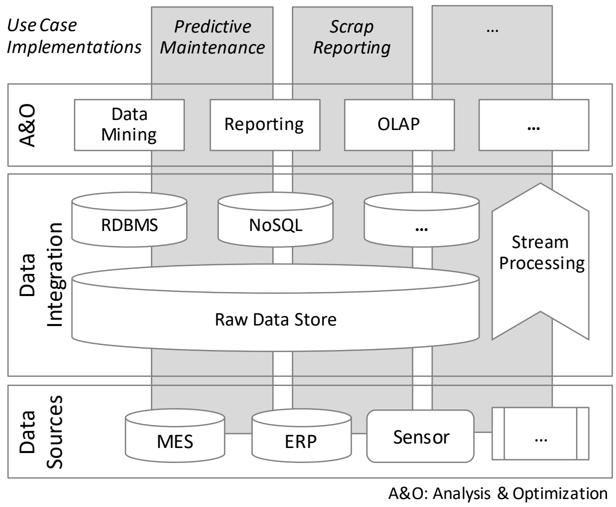

Figure 4: Individual use case implementations on the platform

For each challenge, we describe the practical problem, point out state-of-the-art solutions, describe our own approach and highlight future research directions.

5.1. Overview

The practical challenges we encountered can be structured into three categories: Challenges referring to *analytical solution development*, challenges referring to *employee enablement* as well as challenges referring to *governance.* In the following, we investigate the major challenge in each category:

- Developing standardized and reusable analytical services (see Section 5.2)
- Empowering business domain specialists to do advanced analytics (see Section 5.3)
- Defining a holistic analytics governance (see Section 5.4)

5.2. Analytical Solution Development: Developing Standardized and Reusable Analytical Services

Analytical solution development refers to the implementation of concrete data analytics use cases, e.g., for predictive maintenance or scrap reporting. Our practical experience shows that we typically get specific implementations for each use case and each individual set of data sources (see Figure 4). For instance, predictive maintenance solutions using data mining techniques are implemented from scratch across different factories and manufacturing processes even if they apply to the same type of machine. That is, we significantly suffer from insufficient reusability and portability of analytical solutions across different machines, processes and factories leading to high implementation and maintenance costs.

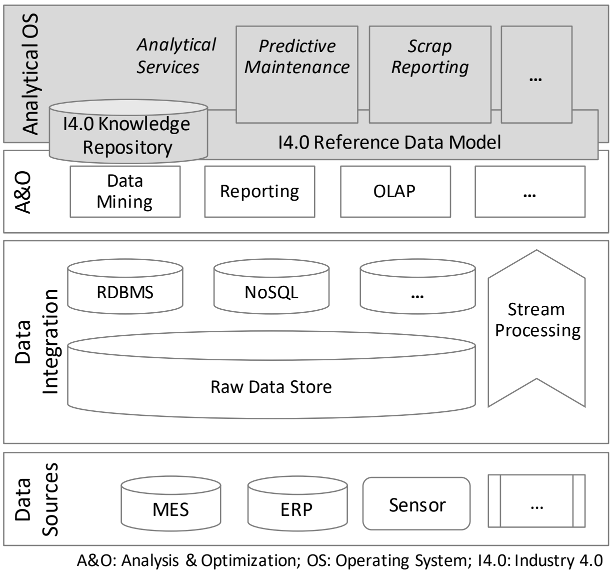

Figure 5: Concept of the analytical operating system

In general, there are two major reasons for this: on the one hand, the use of different tools, e.g., different data mining tools, on the other hand, different data models and data preparation pipelines, even if the source data is of the same structure. Thus, *the challenge is to develop standardized and reusable analytical services which can be parametrized and deployed across different machines, processes and factories*.

To master this challenge, a possible approach would be to use a domain-specific data analytics platform instead of a generic data analytics platform. Domain-specific data analytics platforms, e.g., SAP Predictive Maintenance and Service [38] or GE Predix [12], contain pre-built analytical services for specific use cases. Thus, they enable rapid implementations, however, they are typically proprietary applications hindering the flexible implementation of additional use cases as well as the systematic reuse of certain components, e.g., to reuse selected data preparation steps.

In order to industrialize and speed up analytical solution development, we envision a so-called *analytical operating system* combining the openness and flexibility of a generic data analytics platform with reusable analytical services. As shown in Figure 5, the analytical operating system builds on top of the standardized platform tool stack and abstracts from heterogenous source systems by using an *Industry 4.0 reference data model* to define a common semantic view. Moreover, it stores analytical assets, e.g., single data mining models, data preparation pipelines or reports, in an *Industry 4.0 knowledge repository* and thus enables the development of standardized and reusable *analytical services* which can be deployed across different machines, processes and factories.

In our previous work [16, 17], we designed and prototypically implemented a reference data model for data-driven manufacturing and a corresponding knowledge repository. In order to realize the analytical operating system, comprehensive further research needs to be carried out, especially with respect to:

- A *holistic reference data model for Industry 4.0* which addresses the entire product life cycle, not only the manufacturing phase (see [46] for a survey on existing Industry 4.0 reference models)
- A realization approach to *implement the reference data model in a Lambda-architecture-based data store* and enable flexible integration of structured and unstructured data as well as of batch and streaming data (see [22] for issues on the integration of structured and unstructured data in manufacturing)
- A *standardized and tool-independent meta-model and a corresponding specification format for the definition of modular analytical services* to realize descriptive, diagnostic, predictive and prescriptive analytics (see [34] for the needs for interoperability and standards in data analytics)

5.3. Employee Enablement: Empowering Business Domain Specialists to Do Advanced Analytics

Employee enablement generally refers to empowering end users to utilize the analytics platform. In the course of several advanced analytics projects in manufacturing, e.g., for predictive maintenance or process quality root cause analysis, we experienced a *clash of cultures* between different groups of employees (see Figure 6). These projects are typically organized according to the cross-industry standard process for data mining (CRISP-DM) [19] and require interdisciplinary teams comprising, in particular, business domain specialists and data scientists.

Business domain specialists, e.g., manufacturing process engineers, have comprehensive knowledge about their business domain, its processes and data sources. For instance, they may have detailed know-how on certain machines and manufacturing processes as well as initial ideas for promising data analytics use cases. Yet, they typically have only basic knowledge on data analytics tools and techniques especially regarding advanced analytics. In this paper, we subsume all data analytics techniques beyond classical reporting, dashboarding and online analytical processing under the term advanced analytics, in particular, data mining, machine learning and artificial intelligence techniques (see [4] for a definition of advanced analytics).

Data scientists, e.g., computer scientists or statisticians, have a profound know-how on advanced analytics [28]. They typically have a thorough algorithm and tool

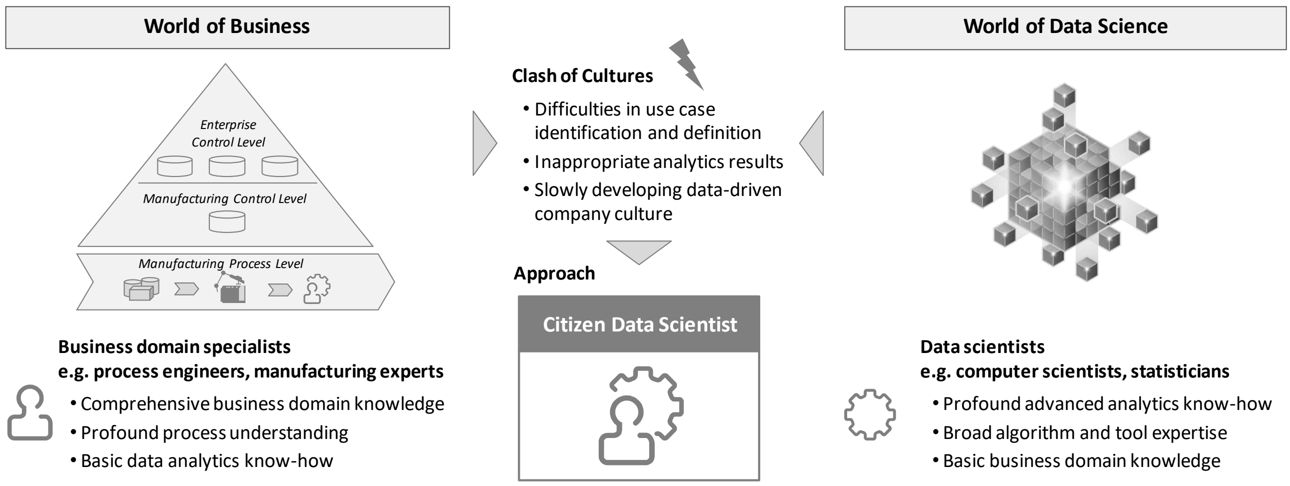

Figure 6: The citizen data scientist approach

expertise for the implementation of advanced analytics use cases, but only basic business domain knowledge.

According to our practical experience, these structural differences between business domain specialists and data scientists frequently cause inefficiency and ineffectiveness in advanced analytics projects and increase the complexity of collaboration. For instance, we often experience difficulties in collaborative use case identification and definition due to different terminologies and educational backgrounds. This leads to imprecise or even inappropriate analytics results, e.g., when business domain specialists assess the results produced by data scientists. Moreover, missing advanced analytics knowledge of business domain specialists prevents data-driven decision making and slows down the development of a data-driven company culture (see [31] for data-driven company culture).

Hence, *the challenge is to empower business domain specialists to do advanced analytics*. To tackle this challenge, the new role of a *citizen data scientist* has recently been proposed [6, 33]. Citizen data scientists combine business domain knowledge with advanced analytics skills in order to bridge the gap between the world of business and the world of data science. That is, their primary job function is outside the field of advanced analytics, as is in the case of business domain specialists, but they are able to produce advanced analytics results. At this, it's important to remark that citizen data scientists do not replace expert data scientists but complement them.

The citizen data scientist approach seems promising to us in order to empower business domain specialists to do advanced analytics and foster data-driven decision making in industrial business processes. However, to the best of our knowledge, there is no sound concept on citizen data scientists yet. We only recognize some initial ideas [8, 33, 44]. Thus, significant further research needs to be done comprising both technical and organizational aspects in order to leverage the role of a citizen data scientist.

Technical aspects refer to *appropriate tools for citizen data scientists* making advanced analytics techniques, especially data mining techniques, usable by non-expert users. Self-service business intelligence tools, e.g., Tableau [42] or Microsoft Power BI [32], represent a starting point by facilitating self-service data visualization for business users. Under the term smart data discovery, these tools now begin to incorporate simple pre-configured data mining techniques [8]. Based on our practical experience, we see a need to combine and enhance features from self-service business intelligence tools with features from graphical data mining tools, e.g., Knime [26] or RapidMiner [36], to facilitate self-service data mining. Moreover, appropriate tools for citizen data scientists should provide domain-specific and reusable analytical services (see Section 5.2), e.g., for data preparation of specific source data or model generation for a specific use case, in order to simplify use case implementation and foster knowledge sharing.

Organizational aspects refer to concepts and methodologies to systematically identify and qualify business domain specialists as citizen data scientists as well as to define their organizational integration. *Qualification* of citizen data scientists particularly requires the development of *interdisciplinary educational plans* combining basic knowledge on data bases, data engineering and statistics with knowledge on advanced analytics algorithms and suitable tools. *Organizational integration* comprises the definition of *collaboration models between expert data scientists and citizen data scientists* especially in large global enterprises such as Bosch. Citizen data scientists are typically located in various business units of an enterprise, whereas expert data scientists form a business-domain-independent center of competence [44]. Thus, *intra-project collaboration* as well as *inter-project collaboration* have to be designed. The former defines how citizen data scientists can be supported by expert data scientists during a project and vice versa. The latter refers to selecting projects or use cases which can completely be done either by citizen data scientists or expert data scientists or which benefit from a joint effort.

5.4. Governance: Defining a Holistic Analytics Governance

In the context of data analytics, *governance* generally refers to establishing and following structures, rules, policies and controls for data analytics activities [2, 10]. This is called *analytics governance* in the following. According to our practical experience at Bosch, there is a huge variety of governance aspects related to an Industry 4.0 analytics platform ranging, e.g., from governance on source data over governance on data analytics projects to governance on data-driven decision making. For example, policies on data ownership have to be developed, appropriate process models for different types of data analytics projects have to be defined and the use of data analytics results in industrial business processes has to be controlled. All these governance aspects significantly determine the effectiveness and the efficiency of data analytics activities.

Hence, *the challenge is to define a holistic analytics governance for an Industry 4.0 analytics platform to ensure its economically beneficial usage*. To master this challenge, a general analytics governance framework is needed which holistically defines governance areas and concepts. This constitutes the starting point for the definition of a concrete analytics governance for the Bosch Industry 4.0 Analytics Platform, which has to be suited to Bosch's organizational structure and strategic goals.

In line with [10], we don't see a sound framework for analytics governance, neither in practice nor in literature. Analytics governance is closely related to and partially overlaps with data governance [25, 40] and both have to be tightly integrated. However, analytics governance goes beyond data governance, e.g., regarding methods and processes for data analytics projects such as CRISP-DM. There are only rudimentary existing works on analytics governance [2, 14, 43]. They underline its importance, name rudimentary governance areas, e.g., people and roles, and discuss guiding principles for the design of a corresponding framework.

Hence, significant further research and practical evaluation have to be carried out in order to define a holistic analytics governance framework. Based on our practical experience at Bosch, we want to highlight the following core issues regarding analytics governance that may guide future research:

- *Balancing trust and flexibility:* As described in Section 4.2, the Bosch Industry 4.0 Analytics Platform provides the whole range of analytical capabilities from descriptive to prescriptive analytics. These different capabilities have totally different requirements regarding trust and flexibility. For instance, classical descriptive reporting requires fully trustworthy results and thus standardized and governed data models, data preparation pipelines and reporting tools to ensure consistent management decisions [23]. In contrast, predictive and prescriptive analytics benefit from flexibility regarding, e.g., ad-hoc integration of new data sources and self-service data exploration with various tools [45]. Hence, a holistic analytics governance has to balance trust and flexibility across the entire analytics platform especially with respect to self-service data preparation versus managed ETL as well as user-driven data discovery versus governed data analytics projects.
- *Integrating with the existing business intelligence and data warehouse platform:* Apart from the Bosch Industry 4.0 Analytics Platform, there is still the existing business intelligence and data warehouse platform at Bosch focusing on reporting and dashboarding on curated enterprise data from various ERP systems. This platform contains valuable enterprise data relevant for data-driven manufacturing. However, it cannot simply be replaced by the analytics platform, because the general requirements and features of a business intelligence and data warehouse platform continue to exist. That means, similar systems would have to be built up as part of the analytics platform considering the huge investments that have already been made in specific data models, data preparation pipelines and reports on the existing business intelligence and data warehouse platform. Therefore, the analytics platform and the business intelligence and data warehouse platform coexist at Bosch. We consider this to be a typical situation in large industrial enterprises having established highly integrated business intelligence and data warehouse platforms which provide source data for an Industry 4.0 analytics platform. Thus, a holistic analytics governance has to define the role of existing business intelligence and data warehouse platforms in the light of an Industry 4.0 analytics platform and detail their integration. In particular, it has to be defined which kinds of use cases are implemented on which platform and how data exchange is governed, for instance, when data warehouse data is transferred to the analytics platform and redundancies as well as potential inconsistencies occur.

- *Managing data-driven decision making:* An Industry 4.0 analytics platform provides the technical basis for data-driven decision making in industrial business processes. The goal is to support or automatize decisions and thus enhance the process [8]. For example in a manufacturing process, quality measurements and corresponding rework decisions can be automated using machine sensor data. For example in a product engineering process, product redesign decisions can be supported using field data. Operational decision support and particularly decision automation cause certain risks, e.g., in case of wrong decisions, and thus need to be managed. Consequently, a holistic analytics governance has to define rules and policies for data-driven decision making including criteria for suitable business processes, appropriate data analytics techniques as well as the degree of human intervention.

6. Summary and Conclusion

This paper is based on our practical experience at Bosch when building an Industry 4.0 analytics platform. The Bosch Industry 4.0 Analytics Platform is designed for the worldwide manufacturing network of Bosch and follows a generic approach to address the entire range of data analytics use cases, from descriptive to prescriptive analytics across both enterprise data and big data. Built upon a Lambda architecture approach and a standardized tool stack, it facilitates the exploitation of the huge amounts of data across the industrial value chain.

Successfully building and establishing such a platform involves various challenges far beyond tools and technologies for data analytics. It is particularly about (1) industrializing analytical solution development on top of the platform, (2) empowering business domain specialists to do advanced analytics and (3) defining a holistic analytics governance to ensure efficiency and effectiveness of the platform's usage.

In conclusion, it can be stated that all these challenges require comprehensive further research especially with respect to (1) the specification of modular and reusable analytical services, (2) appropriate tools and organizational models for citizen data scientists and (3) frameworks for analytics governance. From our point of view, these issues can only be addressed by interdisciplinary research of computer science, manufacturing engineering and information systems in order to leverage data analytics in large industrial enterprises.

Acknowledgments

The author would like to thank Dieter Neumann and Thomas Müller for their valuable comments and continuous support of this work.

References

1. Apache Software Foundation (2017) Apache Hadoop. http://hadoop.apache.org/. Accessed Sept. 2017
2. Avery A, Cheek K (2015) Analytics Governance: Towards a Definition and Framework. In: Proceedings of the 21st Americas Conference on Information Systems (AMCIS) 2015
3. Bauernhansl T (2014) Die Vierte Industrielle Revolution – Der Weg in ein wertschaffendes Produktionsparadigma. In: Bauernhansl T, Hompel Mt, Vogel-Heuser B (eds) Industrie 4.0 in Produktion, Automatisierung und Logistik. Anwendung, Technologien, Migration. Springer Vieweg, Wiesbaden, pp 5–35
4. Bose R (2009) Advanced analytics: opportunities and challenges. Industrial Management & Data Systems 109(2): 155–172
5. Brettel M, Friederichsen N, Keller M et al. (2014) How Virtualization, Decentralization and Network Building Change the Manufacturing Landscape: An Industry 4.0 Perspective. International Journal of Science, Engineering and Technology 8(1): 37–44
6. Burton B, Walker MJ (2015) Gartner Hype Cycle for Emerging Technologies
7. Chamoni P, Gluchowski P (2017) Business Analytics - State of the Art. Controlling & Management Review 61(4): 8–17
8. Davenport TH, Harris JG (2017) Competing on analytics. The new science of winning. Harvard Business Review Press, Boston
9. Dedić N, Stanier C (2017) Towards Differentiating Business Intelligence, Big Data, Data Analytics and Knowledge Discovery. In: Proceedings of the 5th International Conference on ERP Systems, Revised Papers. Springer, Cham, pp 114–122
10. Espinosa JA, Armour F (2016) The Big Data Analytics Gold Rush. A Research Framework for Coordination and Governance. In: Proceedings of the 49th Hawaii International Conference on System Sciences (HICSS) 2016. IEEE, pp 1112–1121
11. Gandomi A, Haider M (2015) Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management 35(2): 137–144
12. GE (2017) Predix Platform. https://www.ge.com/digital/predix/platform. Accessed Sept. 2017
13. Gölzer P, Cato P, Amberg M (2015) Data Processing Requirements of Industry 4.0 - Use Cases for Big Data Applications. In: Proceedings of the 23th European Conference on Information Systems (ECIS), paper 61
14. Grochow J (2012) IT Infrastructure to Support Analytics: Laying the Groundwork for Institutional Analytics. EDUCAUSE Research Bulletin

15. Gröger C (2015) Advanced Manufacturing Analytics - Datengetriebene Optimierung von Fertigungsprozessen. Josef Eul, Lohmar
16. Gröger C, Stach C (2014) The Mobile Manufacturing Dashboard. In: Proceedings of the 2014 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops). IEEE, Piscataway, pp 138–140
17. Gröger C, Schwarz H, Mitschang B (2014) The Manufacturing Knowledge Repository. Consolidating Knowledge to Enable Holistic Process Knowledge Management in Manufacturing. In: Proceedings of the 16th International Conference on Enterprise Information Systems (ICEIS) 2014. SciTePress, pp 39–51
18. Gröger C, Kassner L, Hoos E et al. (2016) The Data-Driven Factory. Leveraging Big Industrial Data for Agile, Learning and Human-Centric Manufacturing. In: Proceedings of the 18th International Conference on Enterprise Information Systems (ICEIS) 2016. SciTePress, pp 40–52
19. Han J, Kamber M, Pei J (2012) Data Mining. Concepts and Techniques, 3rd edn. Morgan Kaufmann, Waltham
20. Jeschke S, Brecher C, Meisen T et al. (2017) Industrial internet of things and cyber manufacturing systems. In: Jeschke S, Brecher C, Song H et al. (eds) Industrial internet of things. Cybermanufacturing systems. Springer, Cham, pp 3–20
21. Kart L, Linden A, Schulte WR (2013) Extend Your Portfolio of Analytics Capabilities. Gartner Research Note G00254653
22. Kassner L, Gröger C, Mitschang B et al. (2014) Product Life Cycle Analytics - Next Generation Data Analytics on Structured and Unstructured Data. Procedia CIRP 33: 35-40
23. Kemper H-G, Baars H, Mehanna W (2010) Business Intelligence, 3rd edn. Vieweg+Teubner, Wiesbaden
24. Kemper H-G, Baars H, Lasi H (2013) An Integrated Business Intelligence Framework. Closing the Gap Between IT Support for Management and for Production. In: Rausch P, Sheta AF, Ayesh A (eds) Business Intelligence and Performance Management. Theory, Systems and Industrial Applications. Springer, London, pp 13–26
25. Khatri V, Brown CV (2010) Designing data governance. Communications of the ACM 53(1): 148–152
26. Knime (2017) Knime Analytics Platform. http://www.knime.com. Accessed Sept. 2017
27. Lee J, Kao H-A, Yang S (2014) Service Innovation and Smart Analytics for Industry 4.0 and Big Data Environment. Procedia CIRP 16: 3–8
28. Loshin D (2013) Big data analytics. Morgan Kaufmann, Amsterdam
29. Lv Z, Song H, Basanta-Val P et al. (2017) Next-Generation Big Data Analytics. State of the Art, Challenges, and Future Research Topics. IEEE Transactions on Industrial Informatics 13(4): 1891–1899
30. Marz N, Warren J (2015) Big data. Principles and best practices of scalable realtime data systems. Manning, New York
31. McAfee A, Brynjolfsson E (2012) Big data: The management revolution. Harvard Business Review 90(10): 60–68
32. Microsoft (2017) Power BI. http://powerbi.microsoft.com/. Accessed Sept. 2017
33. Morgan L (2015) Citizen Data Scientists: 7 Ways To Harness Talent. https://www.informationweek.com/big-data/big-data-analytics/citizen-data-scientists-7-ways-to-harness-talent/d/d-id/1321389. Accessed Sept. 2017
34. OECD (2015) Data-driven innovation. Big data for growth and well-being. OECD, Paris
35. O'Leary DE (2013) Artificial Intelligence and Big Data. IEEE Intelligent Systems 28(2): 96–99
36. RapidMiner (2017) RapidMiner Data Science Platform. http://www.rapidminer.com. Accessed Sept. 2017
37. Robert Bosch GmbH (2017) Bosch Group. http://www.bosch.com/bosch-group/. Accessed Sept. 2017
38. SAP (2017) SAP Predictive Maintenance and Service. https://www.sap.com/products/predictive-maintenance.html. Accessed Sept. 2017
39. Shih W, Ludwig H (2016) The Biggest Challenges of Data-Driven Manufacturing. https://hbr.org/2016/05/the-biggest-challenges-of-data-driven-manufacturing
40. Soares S (2012) Big data governance. An emerging imperative. MC Press, Boise
41. Stark J (2011) Product lifecycle management. 21st century paradigm for product realisation, 2nd edn. Springer, London
42. Tableau (2017) Tableau software for business intelligence and analytics. http://www.tableau.com. Accessed Sept. 2017
43. Tapadinhas J (2016) How to implement a modern business intelligence and analytics platform. Gartner Research Note ID G00291781
44. Thompson JK, Rogers SP (2017) Analytics. How to win with intelligence. Technics Publications, Basking Ridge
45. Topchyan AR (2016) Enabling data driven projects for a modern enterprise. Proceedings of the Institute for System Programming of the RAS 28(3): 209–230
46. Weber C, Königsberger J, Kassner L et al. (2017) M2DDM – A Maturity Model for Data-Driven Manufacturing. Procedia CIRP 63: 173–178