



# A Hybrid Approach to Implement Data Driven Optimization into Production Environments

Rachaa Ghabri<sup>(✉)</sup>, Pascal Hirmer, and Bernhard Mitschang

Institute of Parallel and Distributed Systems, University of Stuttgart,  
Universitätsstr. 38, 70569 Stuttgart, Germany  
{rachaa.ghabri,pascal.hirmer,bernhard.mitschang}@ipvs.uni-stuttgart.de

**Abstract.** The potential of data analytics to improve business processes is commonly recognized. Despite the general enthusiasm, the implementation of data-driven methods in production environments remains low. Although established models, such as CRISP-DM, offer a structured process in order to deploy data analytics in the industry, manufacturing companies still need to choose a starting point, assess the business benefit, and determine a pragmatic course of action. In this paper, we introduce an approach to handle these issues based on a case study from automotive manufacturing. The results are discussed based on a set of requirements derived from the case study.

**Keywords:** Data driven optimization · Production environment  
Top-down · Bottom-up

## 1 Introduction

Globalization, shorter product life cycles, and rapidly changing customer needs lead to increasing competitive pressure in the manufacturing industry [1]. In addition to the high product quality and variety, flexibility and short delivery times are also important success factors [2]. Thus, efficient and continuously improved production processes are key prerequisites for a manufacturing company to become and remain successful in the market [3]. In other business sectors, especially e-commerce and internet based services, big data computing and analytics are successfully used for data driven process optimization [4]. This fact puts forth the potential of data driven optimization as a means to boost business processes. Manufacturing companies can also exploit this potential and use data driven optimization in order to meet the ever increasing demands.

In most cases, companies use the *bottom-up approach*, where business-relevant knowledge is searched in all the available data, for example, by using data mining techniques [5]. This approach is characterized by insufficient focus on specific business objectives and strategies of the respective company, as well as relatively

high investments. Furthermore, the bottom-up approach might be over engineered for the use for production processes, since the semantics and structure of the generated data are usually well known.

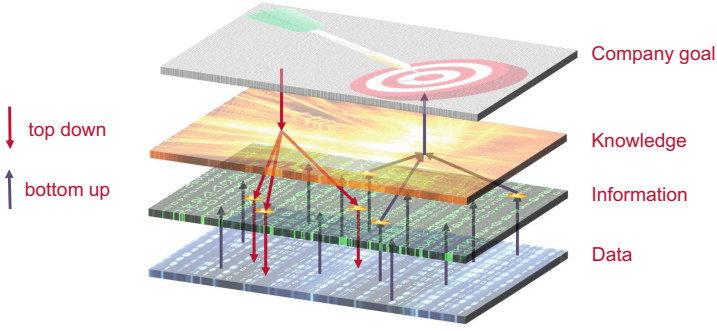
Another well-established alternative is the *top-down approach*, where selective gathering and analysis of data are conducted solely based on specific business goals [6]. This approach bears the risk of missing business relevant knowledge and leaving lucrative optimization potential unlevered.

In both cases, having access to important amounts of data and disposing of powerful IT-Tools do not necessarily lead to a successful data driven optimization. In addition, companies must also embrace an efficient course of action to ensure (i) a sufficient focus on the strategic company goals, while (ii) still leveraging every business potential, and (iii) maintaining a reasonable expense-benefit ratio. For the mentioned reasons, the bottom-up and top-down approaches do not deliver optimal results. To antagonize these problems, there needs to be a strategic course of action, which combines the advantages and avoids the disadvantages of both. In this paper, we introduce our idea of a hybrid approach to achieve maximum benefit out of data-driven optimization and support it with a real-life case study from the production of car electronics. An in-depth description of the associated steps in order to apply the hybrid approach as a methodology will be delivered in a subsequent paper.

The remainder of this paper is structured as follows: Sect. 2 describes the background and related work. In Sect. 3, a case scenario is described, which is used to derive requirements for our approach. Section 4 contains the main contribution of our paper: a hybrid approach to implement data driven optimization into production environments. In Sect. 5, we discuss to what extent the hybrid approach meets the requirements of the case scenario. Finally, Sect. 6 concludes the paper and gives an outlook on future work.

## 2 Background and Related Work

In recent years, procedures have been developed to integrate data-driven optimizations into existing IT and process environments of companies. Usually, these approaches are either classified as bottom-up or top-down. This classification originates from the widely used big data pyramid [7,8], which is depicted in Fig. 1. The bottom layer of the pyramid, the data layer, represents low level data, which can be stored into different, distributed, heterogeneous IT systems or even into so-called data lakes [9,10]. The low level data can then be processed and aggregated, for example, by applying data mining techniques [11], in order to generate information, which is represented through the second layer of the pyramid. This information describes interesting, previously unknown patterns in the data. Interlinking this information and combining it with domain specific expertise leads to the third layer of the pyramid, which represents business relevant knowledge. This knowledge can be used as a basis for further actions in order to reach company goals, for example, by purposefully altering the business processes. A pass through the pyramid in Fig. 1 can be carried out either according to the bottom-up or the top-down approach.



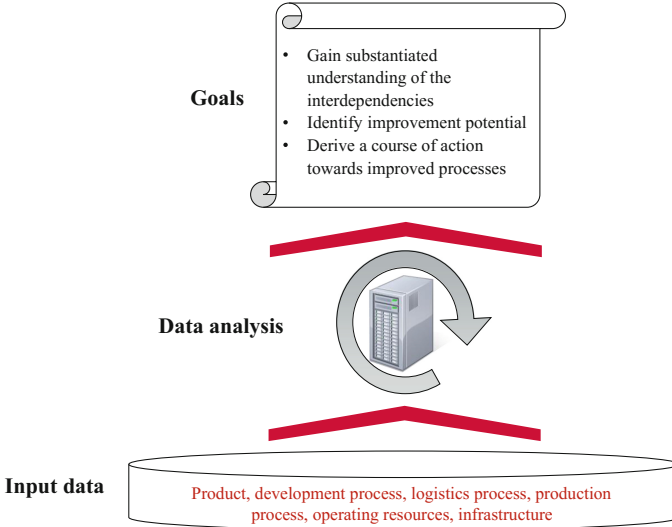
**Fig. 1.** Top-down vs. bottom-up approach

In the bottom-up approach, raw data produced by heterogeneous distributed systems is used as a foundation to derive knowledge that can, for example, lead to the adaptation of processes to increase their efficiency. A precondition for this approach is a holistic, consistent foundation of data to extract or compute information and, consequently, the desired knowledge. For this purpose, data mining techniques can be used in order to recognize interesting patterns in the data. An approved methodology for this is the Knowledge Discovery in Databases (KDD) process as introduced by Fayyad et al. [12]. The bottom-up approach works well in approaches where the data sources and the goals that should be achieved by data analytics are well-known, e.g., when executing previously modeled data flow pipelines [13] or recognizing situations based on context data [14, 15].

However, once a company chooses to apply the bottom-up approach for data driven optimization, it is confronted with a major issue: there is no warranty that all the efforts lead to good results, and, furthermore, there is no reference, which results should be achieved in order to consider the project successful. Furthermore, the recognized patterns in the data may even be misleading, or may be interpreted in a wrong way, leading to no improvements or even to a worsening of the business processes.

On the other hand, the top-down approach builds on specific company goals, which have been derived from a thorough analysis of the enterprise's business processes and IT systems. Based on the specific goals, suitable data and adequate analysis techniques are purposefully selected. The top-down approach is a target-oriented methodology and is more likely to lead to useful results. However, it comes with the risk of missing important information due to the specific, narrow view on the data. Besides, it can be very difficult to decide which data can be considered as relevant to reach the defined goals.

In summary, the bottom-up and top-down approaches have their respective advantages and shortcomings. With the hybrid approach, we aim for in this paper, we use a combination of both approaches to emphasize their advantages and avoid their disadvantages.



**Fig. 2.** Using data analytics in the case scenario (bottom-up approach)

### 3 Case Scenario and Requirements

In this section, we depict the need for a hybrid approach and present one instance of its usage by describing a real-world case scenario from the automotive industry, more specifically from the production of car electronics. Based on this case scenario, we derive a set of requirements for our approach.

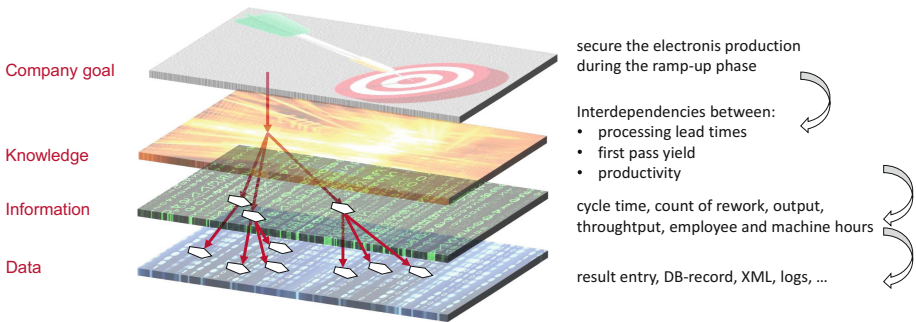
#### 3.1 Case Scenario

The increase in complexity of modern cars electronics in terms of architecture, performance, and communication data is one of the reasons why their production processes become more and more challenging for automotive manufacturers [16]. In this case scenario, a large automotive manufacturer aims for improving its electronics production processes in several manufacturing plants. The electronics production processes are the steps of the final assembly, which consist of mounting all the electronic components of the car, flashing the electronic control units with the customer specific software, calibrating the driver assistance systems, commissioning as well as conducting the functional and final testing for all of the electronic components of the car.

In order to improve this part of the product creation process, it is advisable to consider the preceding steps as well. Thus, we are looking at a process chain, which reaches out from technical development, through production planning, up to the operational production. In this case scenario, the process chain is consistently digital, and therefore, generates large amounts of detailed data. This data is usually decentralized, inhomogeneous, and includes detailed information about the product, the development, logistics and production processes,

as well as the used equipment and infrastructure. By the means of data driven optimizations, the automotive manufacturer seeks to gain deep, numerically supported understanding of the interdependencies within the selected business processes and identify improving potentials, as well as deduce an adequate course of action to exploit this potential. Using data analytics (cf. Fig. 2), the manufacturer aims at reaching optimization goals, e.g., a course of action towards improved processes.

At the beginning of this project, the responsible employees of the car manufacturer are confronted with the task of gaining and keeping a perspective of the large amounts of inhomogeneous, apparently incoherent data. Furthermore, several factors make it difficult to decide, which data sets should be considered significant for the analysis. First, the examined business processes show a lack of transparency due to their high complexity and the numerous interdepartmental interfaces. Secondly, the documentation of the processes and the meta data might show some incoherency. Lastly, the conjunction of data sources is not always conducive, which can undermine data with a potential for business relevant knowledge. These conditions render the data preprocessing and integration a cumbersome task that can also affect the motivation.



**Fig. 3.** Top-down approach for the case scenario

As mentioned in Sect. 2, the bottom-up approach starts with collecting and storing all available data. The architectural and structural components of an adequate IT-solution for this job are usually a huge financial investment, while it is not clear, whether process improvements would occur and if they would account for company goals with high priority. This uncertainty makes the budget clearance for such projects more difficult. Additionally, the company would have to commit to a certain IT-solution prior to conducting a spike test to ensure the feasibility and the suitability. Here lies a risk that the IT-solution would emerge as unfit for the business and operating environment of the company [17].

Applying the top-down approach in this case scenario would start with defining a specific company goal, e.g. securing the electronics production during the ramp-up phase of a new car model. Based on this given goal, the necessary

knowledge needs to be determined. For instance, the interdependencies between the processing lead times, the first pass yield, and the productivity of the whole production plant or of a single work station should be useful. The next step is to specify the information, which would lead to the needed knowledge. In this case, the information about the cycle time of a single operation, the count of rework and mistakes, the output, the throughput and the number of employee and machine-hours should be considered. Up to this point of the data driven optimization, it is irrelevant, which IT-tools and architectural components will be used. For the transition from the information layer to the data layer it is, however, necessary to determine the right data sources, design the information model including metadata management, as well as to conduct the adequate data processing. Therefore, it is necessary to make a decision about the IT-solution to be employed. Using the top-down approach for this example is depicted in Fig. 3.

While the top-down approach is more likely to help the company reach its optimization goals, its scope of action is limited to one single issue. Thus, using this approach only allows a parochial view of the data and the improvement potentials, rather than considering the wider context. For instance, when the top-down approach is used in the example above, important insights, such as the impact of the infrastructure on the processing lead time, would remain undetected. Besides, by focusing on a given goal, the domain experts in the company miss out on an opportunity to expand their domain knowledge and discover previously unknown interdependencies within the process chain.

In summary, both approaches cannot provide a satisfying solution for the usage of data analytics in order to improve production processes of this or other scenarios. In the following section, we derive a set of requirements to cope with the mentioned issues, which build the foundation of our approach.

### 3.2 Requirements

The hybrid approach, we aim for in this paper, minimizes the risks and combines the advantages of the approaches described above. We define the following requirements for our approach:

- (R1) Contribution to high-priority company goals: The hybrid approach needs to ensure that the data driven optimization is set up to contribute to strategic, highly prioritized goals of the company. Thus, the first step of the approach must consist in defining a concrete outcome of the project. By doing so, it is possible to evaluate and rank a specific data driven project based on the company's current priorities.
- (R2) Full development of the potential for improvement: The hybrid approach must ensure that the data analysis reveals every worthy room for improvement: as a counterpart for the pragmatic implementation, the long-term expectation out of data driven optimization is to look into every potentially value-adding insight.

- (R3) Optimal cost-benefit ratio: The hybrid approach aims to achieve an optimal cost-benefit ratio out of data driven optimization: it avoids investments with a long payback period. Instead, it relies on incremental investments with many “low-hanging fruits”.
- (R4) Promotion of feasibility: The approach must promote the feasibility of data driven optimization within the business- and operating-guidelines of the company: data driven optimization is not conducted for its own sake, but rather to bring a practical benefit for the company. Therefore, they should be conducted pragmatically and with minimum distortion of the core business.

## 4 Hybrid Approach for Data Driven Optimization

The goal of this paper is an approach to implement data driven optimization into production environments, while minimizing the disadvantages and highlighting the advantages of the established bottom-up and top-down approaches. The hybrid approach consists of a purposeful, structured alteration and combination of the top-down and bottom-up approaches in order to join a motivating effectiveness with a holistic performance, and at the same time to avoid high, uncertain investment. Initially, a set of use cases are concluded in the style of the top-down approach. In addition to fulfilling the specific purpose of the use case, each successful execution will reveal business-valuable data sets. These are the data sets which evidentially lead to profitable knowledge for the company. Such data sets are referred to as *data treasures* in the context of this paper. Using the bottom-up approach, the data treasures are then analyzed and the contained information is correlated in order to gain insights beyond the discrete use cases. In doing so, companies can ensure a maximum benefit out of data-driven optimizations while holding the risks at a viable level. Figure 4 shows the steps of the hybrid method, which is explained in the following sections.

### 4.1 Derive, Prioritize and Execute Use Cases

The first phase of the hybrid approach is based on the top-down approach. As mentioned in Sect. 2, this approach begins with the definition of a business goal that is in line with the company’s strategic objectives. Queries with a direct reference to the production field are then derived from the business goal. The queries should be formulated as precisely as possible, and the corresponding frame conditions should be specified in order to answer them in the context of specific use cases by using analytics techniques. We suggest the following references in order to convert a query into a manageable use case:

- Which type of analytics, i.e., descriptive, diagnostic, predictive, or prescriptive [18], is suitable to answer the respective query?
- What are the key performance indicators, parameters and influencing factors involved in the query?
- Is the underlying data already available and, if not, what needs to be done to make it available?

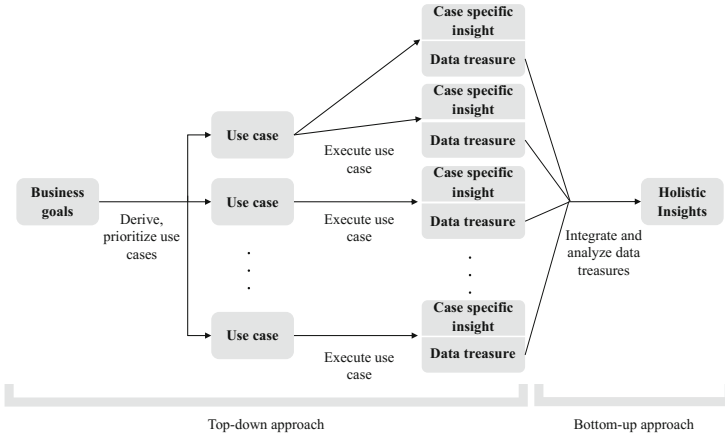


Fig. 4. Steps of the hybrid approach

- Which are the sources of the underlying data and which format does the raw data have?
- Which requirements must the data processing meet, e.g., real-time or incremental processing?
- How long is the period of time that is considered in the analysis and how frequently will the analysis be conducted?
- What practical benefit for the company comes with answering the query?

After converting a query into a use case and based on the answers to the questions above, a potential analysis is to be conducted. At this point, we recommend to look into the following features to assess the priority of a given use case: (i) acuteness, defining to what extent the use case attends to urgent issues of the company, (ii) feasibility, describing how much effort goes into providing and processing the needed data, and (iii) relevance, examining the bearing of the benefit. The potential analysis helps the company identify result-oriented, data-based use cases in an efficient, structured and repeatable manner.

For the highly-prioritized use cases, the required data sources are made available, access authorization is managed, and data security measures are taken. Afterwards, the data is processed, e.g., through validation, cleaning, and aggregation, in order to prepare it for the subsequent analysis. In the analysis step, statistical evaluation is used in order to answer the query with the help of the data. The results of the analysis are then made comprehensible by means of appropriate visualization. The latter is then evaluated by the domain experts and used as support to conclude a course of actions. The sequence of the first phase of the hybrid approach is depicted in Fig. 5.

The first phase of the hybrid approach are conducted in a cyclical manner. This means that the output of a successfully executed use case, i.e., the gained in-sights, may influence the input of the next use case, for example, through the adaptation of a defined business goal or the definition of new ones. The





**Fig. 5.** The first phase of the hybrid approach (Steps in red color are conducted by the business department, steps in grey color by the IT department) (Color figure online)

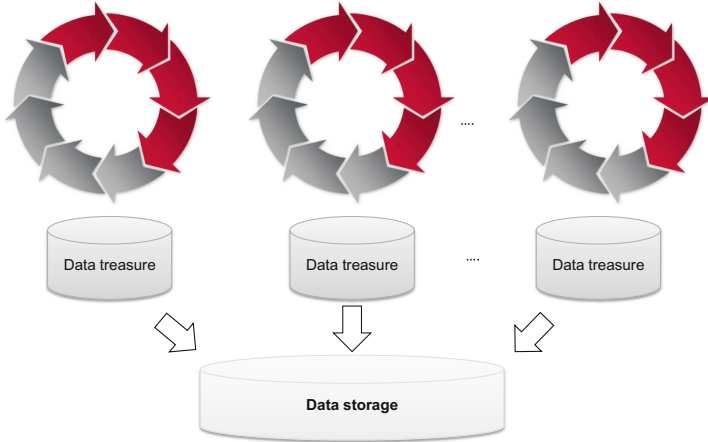
execution of the first phase of the hybrid approach calls for the collaboration of the respective business departments and the IT department. In Fig. 5, the steps marked in red color are to be conducted by the business department, while the steps marked in grey color are the tasks of the IT department. To reach maximum benefit, it is recommended to comply with the allocation of tasks, so that each department can concentrate on its core expertise.

## 4.2 Integrate and Analyze Data Treasures

The executed use cases conduce to identify the parts of the data jungle, which contain information with business value. As already mentioned, this data shall be referred to as data treasure. Once a data treasure has been identified, it is made available in a central data storage, for example, a data lake, (cf. Fig. 6). This way, the central data storage, i.e. the data lake, will only contain data with confirmed usefulness, and will expand with every conducted use case. The data treasures of a specific use case, which show mutual correlation, are then assigned to one cluster and should be considered as a coherent entity.

## 4.3 Collecting and Accessing Data Treasures

In the style of the bottom-up approach, the established entities are then examined in order to find correlations with each other or with further parameters from different yet related use cases. By doing so, the analysis is carried out not only within the boundaries of single use cases, but rather on a holistic level. Since



**Fig. 6.** Collecting and accessing data treasures

this step is likely to be sophisticated and costly, it should be ensured that the efforts are well-invested. For that reason, the risk of trailing away with irrelevant or pseudo-correlations needs to be minimized. We recommend this step of the hybrid approach to adhere specific frame conditions in order to maintain the efficiency. For instance, integrating and analyzing the data treasures can be carried out for a specific period of time, a specific car model, or a specific manufacturing technology.

## 5 Discussion

In this section, we discuss our approach in terms of fulfilling the requirements from Sect. 3.2. One of the main features of the hybrid approach is its pronounced goal-orientation. The first phase of the hybrid approach ensures that careful consideration is dedicated to defining and selecting project objectives that are in line with the company's goals in order to stay focused on what is strategically important; hence, the first requirement (R1), i.e. contribution to high-priority company goals, is fulfilled. Nevertheless, the hybrid approach allows for exploiting the potential of data analysis beyond rigidly set objectives. Companies can reach high-level value through the purposeful application of the bottom-up approach, which makes sure that none of the potentials for data-driven optimization remain undiscovered. Therefore the hybrid approach also meets the requirement of allowing a full development of the potential for improvement (R2). Besides, as the phases of the hybrid approach are meant to be executed consecutively, the company will have the possibility to gradually ascertain the true business value of the available data sources; to concentrate on utilizing data analysis as a means to improve the business processes. Since the hybrid approach initially relies on the consecutive implementation of several stand-alone use cases, it does

not call for a primary large investment. It rather favors gradual investments with perceptible impact. Furthermore, the company is able to avoid committing to a costly, sophisticated IT-solutions before thoroughly investigating the specific circumstances. For these reasons the hybrid approach is in line with the requirement (R3) of achieving an optimal cost-benefit ratio. In terms of the requirement of promoting the feasibility (R4), the hybrid approach is characterized by the sensible, practical usage of data analysis in production environments. Due to the sequence of its phases, the hybrid approach provides the company with the opportunity to readjust its course of action in the manner of a loop control-system. Moreover, the design of the hybrid approach allows to achieve quick-wins, which accounts for a sense of achievement among the involved employees and results in a higher motivation.

## 6 Conclusion and Future Work

Data driven optimization is an effective, innovative method for revealing interdependencies and detecting anomalies within the production processes, in order to make them more transparent, stable and controllable. However, a pragmatic, goal oriented and yet holistic approach is key to deploy the full potential of this method. This can be accomplished by adapting and combining the top-down and bottom-up approach. In this paper, we explained the potential analysis of data driven optimization in the production environment and introduced our idea for a hybrid approach for implementing it. In future work, we will deliver further details of the concepts as well as introduce an in-depth method to apply them. For instance, we will depict applicable approaches integration approaches and techniques, in order to interlink the case-specific data-treasures. Furthermore, we intend to look into quantifying the advantage of the hybrid approach in comparison with the conventional top-down and bottom-up approaches in terms of explicit figures.

## References

1. Khan, A., Turowski, K.: A survey of current challenges in manufacturing industry and preparation for industry 4.0. In: Abraham, A., Kovalev, S., Tarassov, V., Snášel, V. (eds.) Proceedings of the First International Scientific Conference “Intelligent Information Technologies for Industry” (IITI 2016). AISC, vol. 450, pp. 15–26. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-33609-1\\_2](https://doi.org/10.1007/978-3-319-33609-1_2)
2. Jacob, F., Strube, G.: Why go global? The multinational imperative. In: Abele, E., Meyer, T., Näher, U., Strube, G., Sykes, R. (eds.) Global Production, pp. 2–33. Springer, Heidelberg (2008). [https://doi.org/10.1007/978-3-540-71653-2\\_1](https://doi.org/10.1007/978-3-540-71653-2_1)
3. Slack, N., Chambers, S., Johnston, R.: Operations Management. Pearson Education, New York (2010)
4. Bryant, R., Katz, R.H., Lazowska, E.D.: Big-data computing: creating revolutionary breakthroughs in commerce, science and society (2008)
5. Raval, K.M.: Data mining techniques. Int. J. Adv. Res. Comput. Sci. Softw. Eng. **2**(10) (2012)

6. Couldry, N., Powell, A.: Big data from the bottom up. *Big Data Soc.* **1**(2), 2053951714539277 (2014)
7. Lovelace, R.: The data revolution: big data, open data, data infrastructures and their consequences, by rob kitchin. 2014. Thousand Oaks, California: Sage Publications. 222+XVII. ISBN: 978-1446287484. *J. Reg. Sci.* **56**(4), 722–723 (2016)
8. McAfee, A., Brynjolfsson, E., Davenport, T.H., Patil, D., Barton, D.: Big data: the management revolution. *Harvard Bus. Rev.* **90**(10), 61–67 (2012)
9. Gartner: Data lake. Gartner IT Glossary (2017)
10. Held, J.: Will data lakes turn into data swamps or data reservoirs? (2014)
11. Han, J., Kamber, M., Pei, J.: Mining frequent patterns, associations, and correlations. In: *Data Mining: Concepts and Techniques*, 2nd edn., pp. 227–283. Morgan Kaufmann Publishers, San Francisco (2006)
12. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: The KDD process for extracting useful knowledge from volumes of data. *Commun. ACM* **39**(11), 27–34 (1996)
13. Hirmer, P., Behringer, M.: FlexMash 2.0 - flexible modeling and execution of data Mashups. In: Daniel, F., Gaedke, M. (eds.) *RMC 2016. CCIS*, vol. 696, pp. 10–29. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-53174-8\\_2](https://doi.org/10.1007/978-3-319-53174-8_2)
14. Hirmer, P., Wieland, M., Schwarz, H., Mitschang, B., Breitenbücher, U., Sáez, S.G., Leymann, F.: Situation recognition and handling based on executing situation templates and situation-aware workflows. *Computing* **99**, 163–181 (2017)
15. Wieland, M., Hirmer, P., Steimle, F., Gröger, C., Mitschang, B., Rehder, E., Lucke, D., Rahman, O.A., Bauernhansl, T.: Towards a rule-based manufacturing integration assistant. In: Westkämper, E., Bauernhansl, T. (eds.) *Proceedings of the 49th CIRP Conference on Manufacturing Systems (CIRP-CMS 2016)*, Stuttgart, Germany, 25–27 May 2016, *Procedia CIRP*, vol. 57, pp. 213–218. Elsevier, January 2017
16. Hermann, M., Pentek, T., Otto, B.: Design principles for industrie 4.0 scenarios. In: *2016 49th Hawaii International Conference on System Sciences (HICSS)*, pp. 3928–3937, January 2016
17. Tönne, A.: Big Data is no longer equivalent to Hadoop in the industry. In: *Proceedings of 17. Datenbanksysteme für Business, Technologie und Web (BTW)* (2017)
18. Hagerty, J.: 2017 planning guide for data and analytics (2016)