Towards a Decision Support System for Big Data Projects

Matthias Volk, Daniel Staegemann, Sascha Bosse, Abdulrahman Nahhas and Klaus Turowski

Otto-von-Guericke-University, MRCC VLBA, Magdeburg, Germany
{matthias.volk,daniel.staegemann,sascha.bosse, abdulrahman.nahhas,klaus.turowski}@ovgu.de

Abstract. Big data has proved to be one of the most promising trends in recent years. However, many challenges and barriers still exist, especially when it comes to the strategic planning and realization of those kinds of projects. Most of all, the selection and combination of the domain–related technologies represents a sophisticated endeavor that increases the complexity of creating a big data system. Hence, it is not surprising that the demand for experts in this area is steadily increasing. To overcome this problem and the related shortage of required knowledge, in the following paper the concept of a decision support system for the selection of appropriate big data technologies is introduced, in order to implement a given project. Through the use of the design science research methodology a preliminary artifact was developed that provides sophisticated recommendations as well as architectural models and blank systems to support the systems engineering procedure.

Keywords: Big Data, Technologies, Decision Support, System, Design Science

1 Introduction

The amount of data produced by humanity is increasing continuously [1]. This challenge exceeds the capabilities of traditional approaches in different ways, requiring new solutions for this task [2]. As a result, the term big data and its accompanying technologies have emerged. For now, a four-digit range of those solutions are existing, and the number is constantly increasing [3]. This circumstance reinforces the already existing complexity of the selection, combination, and governance in different ways [4]. In contrast to conventional information technology (IT) projects, the efficient handling of the data, in terms of storing and processing, requires the combination of specific *immature* technologies [5]. Reinforced by the prevailing lack of specialists [6] it becomes apparent why the realization of big data projects and the choice of the appropriate technologies represent such a sophisticated endeavor. A possible solution for this task lies in the creation of a decision support system (DSS) for the technology selection in projects [7]. Although decision support systems are common in many domains, such as business [8] or transportation [9], there is to our knowledge no existing approach for big data that solves the

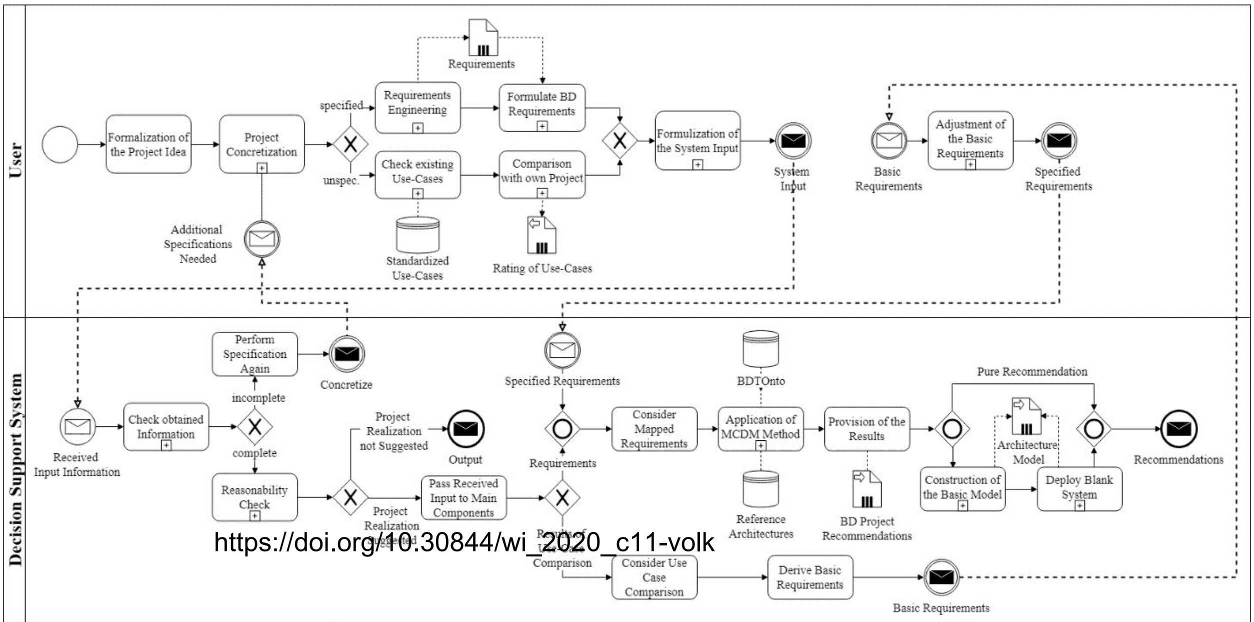
15th International Conference on Wirtschaftsinformatik,
March 08-11, 2020, Potsdam, Germany

https://doi.org/10.30844/wi_2020_c11-volk

aforementioned problems. This leads to the following research question (RQ): *How can decision support for the technology selection in big data projects be facilitated in a systematic and computer-supported way?”.* To answer this, the design science research (DSR) methodology according to Hevner et al. [10] and Peffers et al. [11] are used. Additionally to that, for the creation of the actual concept, the system development research method by Nunamaker et al. [12] is followed as well as various guidelines for the creation of DSS [13–15].

2 Related Work

When it comes to the realization of big data projects, there is a multitude of approaches that try to support the procedure from various perspectives. This becomes apparent, when searching for domain-related phrases in concatenation with big data, such as “project, requirements, use case, decision support, engineering, architecture”. Apart from general descriptions and guidelines [5, 16–20] also technical specifications, such as reference architectures [21–24] and use case descriptions, are available [25–28]. However, to our knowledge, no systematic approach covering the realization from an end-to-end perspective exists. Although the idea of decision support systems in conjunction with big data is not new, many of them are predominately harnessing it for the enhancement of decision support [9, 29–32]. In any case, the starting point is the definition of a project goal and, thus, the requirements that need to be developed beforehand. This is also explicitly highlighted in the contribution by Volk et al. [5] who attempt to present a general workflow for the execution of big data projects including also a general reasonability check [5, 16]. Another approach, presented by Portela et al. [17], seeks to provide a similar solution. Consequently, due to the complexity of this domain, a prior reasonability check may provide initial decision support and represents a first assessment of the planned project [16, 17]. To specify this decision support in a more detailed way, further information about the planned project as well as a sophisticated inference engine are needed. A foundation for both is introduced in [20]. Here, an analytical hierarchy process (AHP) for the selection of the most appropriate technologies is implemented. In general, multi-criteria decision-making (MCDM) methods, are today frequently utilized within DSS [33]. However, this is not necessarily needed, as [19] show. Further, in both cases only a limited amount of tools and technologies are considered. Due to the currently existing amount of those, their interconnections and the specific constraints that have to be regarded, a comprehensive knowledge base is needed. In the context of this, only one particular contribution was found that tries to overcome



https://doi.org/10.30844/wi_2020_c11-volk

the complexity of the sheer amount of existing tools and technologies in this area. In [4], the ontology *BDTOnto* for the classification of big data tools and technologies was introduced. The basic structure is oriented on the crucial steps of a big data project and the performed operations in each phase. Hence, this approach appears to be sufficient as a foundation for such kind of a system. Eventually, through the adaption and extension of the ontology, a variety of different outputs might be imaginable. Additionally to the general recommendations, also concrete artifacts in the form of a diagram or blank systems could be facilitated.

3 The Concept for a Decision Support in Big Data Projects

Based on the findings of the previous investigations, a systematic approach to develop such kind of a solution is conducted. Before the actual conceptualization of the system is taking place, basic requirements need to be developed for the intended solution itself [12]. In order to obtain a better overview of the system, its functionalities, and the interaction with possible stakeholders a use case diagram was modeled (cf. **Figure 1**).

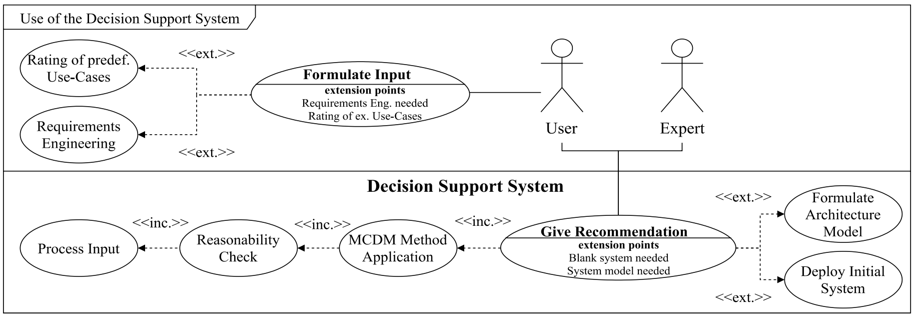


Figure 1. Use case diagram of the intended system

The diagram depicts the intended solution as well as different actors. In addition to the person responsible for the interaction with the system (user), the experts represent the required knowledge necessary for the realization of the proposed recommendations. In terms of this, a multitude of additional information, such as system requirements, the system environment, monetary expenditures for software licenses and others are imaginable. To give the recommendation, the system initially processes the input information from the user. Depending on the time spent for the actual planning, this may come from a thoroughly performed requirements engineering procedure (RE) or the extraction from standardized use-cases which fit the roughly planned project needs (comparison). Once the necessary information is provided, a general reasonability check should be performed, such as described by [16, 17]. This may allow responsible persons to quickly check the meaningfulness before the detailed planning is taking place and therefore have validation for the project itself. For this purpose, adequate methods and a comprehensive knowledge

https://doi.org/10.30844/wi_2020_c11-volk

base are required, especially for the following deduction of concrete recommendations. While suitable ideas for the knowledge base were addressed by various authors, such as in [4, 19], less effort was determined for the latter [20]. However, it was noticed that in any case the observation of multiple criteria appears to be promising and, thus, appropriate methods for MCDM should be applied, as previously highlighted. Apart from the general provision of various recommendations, also a presentation of the particular architecture model or even the deployment of a blank system appears to be desirable. While the first may contain concrete implementation details, the second could deliver an initial setup, to be used for testing and further configuration. Essentially the depicted setup (cf. Figure 1) conforms to the recommended structure of a DSS, highlighting the main parts: inputs, processes, and outputs [13]. However, to represent the "flow of information from the output component to the decision-maker concerning the system's output or performance" [13] also a *feedback* part is needed. This could be realized through the use of a graphical user interface. Whenever inputs (initial requirements, use case comparisons) or adjustments (specifications) are needed, the user will be able to conduct those. After the input is defined, a MCDM method applied and the recommendations given, the user interacts with the system to proceed with the construction of concrete architecture models and the deployment of a blank system originating from the deduced information. Corresponding to the presentation and description of the aforementioned use case diagram, the following functional requirements for the planned solution are derived (cf. **Table 1**).

Table 1. Derived basic requirements of the system

No.	Basic functional requirements
1	Provide recommendations to the user that support the system engineering.
2	Process the received input information from the RE procedure or comparisons.
3	Provide a first reasonability check that identifies the general meaningfulness.
4	Allow adjustments of intermediate results by responsible users.
5	Deploy an initial system that is based on the given recommendation.
6	Formulate an architecture model that is based on the given recommendation.

Before the actual development of the solution is taking place, first, the concept of the underlying knowledge base needs to be defined. As described in section two, a suitable solution for this was already realized in [4]. The developed ontology shall serve as a base for further use and extension of required information, and, thus be adopted for the intended DSS. However, modifications required for the application of the MCDM are needed. This includes, for instance, the attainability of requested project requirements on a technological level or compatibilities to be recognized here. Secondly, the structure of the system is needed. The realization is planned in the form of a webapplication, not only facilitating a greater user experience by using a appealing state-of-the-art framework for the GUI, but also to reduce the effort for future maintenance and updates. Eventually, to ensure that the artifact offers all

https://doi.org/10.30844/wi_2020_c11-volk

functionalities and fulfills the designated purpose, the evaluation pattern by Sonnenberg and vom Brocke [34] will be applied. In doing so, additional to the evaluation of the meaningfulness, feasibility, usability, and functionality, also a large-scale implementation in a business context is planned.

4 Conclusion

Big data proved to be an important asset in many application areas. To ensure that this success continues in the future, new approaches are required to overcome the increasing complexity of big data project realization. Hence, in this work the concept of a decision support system was introduced, providing a first blueprint for the future development of a prototype. This shall not only deliver sophisticated recommendations but also allows the conception of architectural models and the deployment of blank systems.

References

1. Yin, S., Kaynak, O.: Big Data for Modern Industry: Challenges and Trends [Point of View]. Proc. IEEE 103, 143–146 (2015)
2. Chen, C.L.P., Chun-Yang, Z.: Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information Sciences 275, 314–347 (2014)
3. Turck, M. and Obayomi, D.: The Big Data Landscape, http://dfkoz.com/big-data-landscape/
4. Volk, M., Pohl, M., Turowski, K.: Classifying Big Data Technologies - An Ontology-based Approach. In: 24rd Americas Conference on Information 2018. AIS, New Orleans (2018)
5. Volk, M., Jamous, N., Turowski, K.: Ask the Right Questions - Requirements Engineering for the Execution of Big Data Projects. In: 23rd Americas Conference on Information Systems. AIS, Boston (2017)
6. Alharthi, A., Krotov, V., Bowman, M.: Addressing barriers to big data. Business Horizons 60, 285–292 (2017)
7. Volk, M., Staegemann, D., Pohl, M., Turowski, K.: Challenging Big Data Engineering: Positioning of Current and Future Development. In: Proceedings of 4th International Conference on Internet of Things, Big Data and Security, pp. 351–358. SCITEPRESS - Science and Technology Publications
8. Lau, R.Y.K., Liao, S.S.Y., Wong, K.F., Chiu, D.K.W.: Web 2.0 Environmental Scanning and Adaptive Decision Support for Business Mergers and Acquisitions. MIS Quaterly 36, 1239–1268 (2012)
9. Lee, H., Aydin, N., Choi, Y., Lekhavat, S., Irani, Z.: A decision support system for vessel speed decision in maritime logistics using weather archive big data. Computers & Operations Research 98, 330–342 (2018)
10. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design Science in Information Systems Research. MIS Quaterly 28, 75–105 (2004)

https://doi.org/10.30844/wi_2020_c11-volk

11. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A design science research methodology for information systems research. Journal of management information systems 24, 45–77 (2007)
12. Nunamaker, J.F., Chen, M., Purdin, T.D.M.: Systems Development in Information Systems Research. Journal of management information systems 7, 89–106 (1990)
13. Turban, E., Aronson, J.E., Liang, T.-P.: Decision support systems and intelligent systems. Prentice-Hall; Pearson, Upper Saddle River (op. 2005)
14. Power, D.J.: Decision support systems. Concepts and resources for managers. Quorum Books, Westport, Conn. (2002)
15. Shim, J.P., Warkentin, M., Courtney, J.F., Power, D.J., Sharda, R., Carlsson, C.: Past, present, and future of decision support technology. Decision Support Systems 33, 111–126 (2002)
16. Volk, M., Hart, S.W., Bosse, S., Turowski, K.: How much is Big Data? A Classification Framework for IT Projects and Technologies. In: 22nd Americas Conference on Information Systems, AMCIS 2016, San Diego, CA, USA, August 11-14, 2016. Association for Information Systems (2016)
17. Portela, F., Lima, L., Santos, M.F.: Why Big Data? Towards a Project Assessment Framework. Procedia Computer Science 98, 604–609 (2016)
18. Fekete, D., Vossen, G.: The GOBIA Method: Towards Goal-Oriented Business Intelligence Architectures (2015)
19. Lehmann, D., Fekete, D., Vossen, G.: Technology selection for big data and analytical applications. ERCIS - European Research Center for Information Systems, Münster (2016)
20. Lněnička, M.: AHP Model for the Big Data Analytics Platform Selection. Acta Informatica Pragensia 4, 108–121 (2015)
21. Pääkkönen, P., Pakkala, D.: Reference Architecture and Classification of Technologies, Products and Services for Big Data Systems. Big Data Research 2, 166–186 (2015)
22. Kreps, J.: Questioning the Lambda Architecture. The Lambda Architecture has its merits, but alternatives are worth exploring, https://www.oreilly.com/ideas/questioning-the-lambda-architecture
23. Marz, N.: How to beat the CAP theorem, http://nathanmarz.com/blog/how-to-beat-the-cap-theorem.html
24. Martínez-Prieto, M.A., Cuesta, C.E., Arias, M., Fernández, J.D.: The Solid architecture for real-time management of big semantic data. Future Generation Computer Systems 47, 62–79 (2015)
25. Avvenuti, M., Cresci, S., Del Vigna, F., Fagni, T., Tesconi, M.: CrisMap: a Big Data Crisis Mapping System Based on Damage Detection and Geoparsing. Inf Syst Front 20, 993–1011 (2018)
26. Huang, Q., Cervone, G., Jing, D., Chang, C.: DisasterMapper. In: Chandola, V., Vatsavai, R.R. (eds.) BigSpatial 2015, pp. 1–6. The Association for Computing Machinery, Inc, New York, NY (2015)
27. Ta-Shma, P., Akbar, A., Gerson-Golan, G., Hadash, G., Carrez, F., Moessner, K.: An Ingestion and Analytics Architecture for IoT Applied to Smart City Use Cases. IEEE Internet Things J. 5, 765–774 (2018)

https://doi.org/10.30844/wi_2020_c11-volk

28. NIST Big Data Public Working Group: NIST Big Data Interoperability Framework: Volume 3, Use Cases and General Requirements Version 2. National Institute of Standards and Technology, Gaithersburg, MD (2018)
29. Delibašić, B. (ed.): Decision Support Systems V - big data analytics for decision making. Springer International Publishing, Cham (2015)
30. Elgendy, N., Elragal, A.: Big Data Analytics in Support of the Decision Making Process. Procedia Computer Science 100, 1071–1084 (2016)
31. Aversa, P., Cabantous, L., Haefliger, S.: When decision support systems fail: Insights for strategic information systems from Formula 1. The Journal of Strategic Information Systems 27, 221–236 (2018)
32. Poleto, T., Carvalho, V.D.H. de, Costa, A.P.C.S.: The Roles of Big Data in the Decision-Support Process: An Empirical Investigation. In: Delibašić, B. (ed.) Decision Support Systems V - big data analytics for decision making, 216, pp. 10–21. Springer International Publishing, Cham (2015)
33. Razmak, J., Aouni, B.: Decision Support System and Multi-Criteria Decision Aid: A State of the Art and Perspectives. J. Multi-Crit. Decis. Anal. 22, 101–117 (2015)
34. Sonnenberg, C., Vom Brocke, J.: Evaluations in the Science of the Artificial - Reconsidering the Build-Evaluate Pattern in Design Science Research. In: Peffers, K., Rothenberger, M., Kuechler, B. (eds.) Design Science Research in Information Systems. Advances in Theory and Practice, pp. 381–397. Springer Berlin Heidelberg, Berlin, Heidelberg (2012)

https://doi.org/10.30844/wi_2020_c11-volk