Ask the Right Questions: Requirements Engineering for the Execution of Big Data Projects

Full Paper

Matthias Volk, Naoum Jamous, Klaus Turowski
Magdeburg Research and Competence Cluster Very Large Business Applications
Faculty of Computer Science, Otto von Guericke University Magdeburg
{matthias.volk, naoum.jamous, klaus.turowski}@ovgu.de

Abstract

Big data and data analytics related projects are attracting researchers, practitioners, and policymakers around the globe. In order to increase the probability of a project success and to get competitive advantages, stakeholders need to have solid understanding of the main objectives, risks, requirements, and the used technologies. This is a challenging task, especially for a big sized project based on new introduced technologies. Using the design science methodology, this paper attempts to answer the following question: How can the requirements of a project be identified and analyzed to determine the suitability of a combined big data technology application? The main objective is a suitable procedure to determine the most relevant requirements as well as the suitability of the combined use of big data technologies in projects. Thus, a new process model will be proposed, detailed, and evaluated. Then, all findings will be discussed and a comprehensive outlook for further research will be highlighted.

Keywords

Big Data, Big Data Projects, Requirements Engineering, IT Project Success

Introduction

Since the first introduction of Information Technology (IT) in the middle of the last century, IT usage in enterprises has dramatically evolved, especially in the last 30 years. Today IT is one of the main pillars in global competition. Technologies such as Internet, e-commerce, Internet of Things (IoT), Cloud computing, social media, and big data have led to the exponential growth of data and to the complex, heterogeneous and dynamic IT system landscapes that can support business processes and value chains of several enterprises. In the 2014 Gartner Hype Cycle, big data was to move into the trough of disillusionment. In the Hype Cycle of 2015, only specific big data issues were included and no longer the buzzword itself (Peddibhotla 2015). Big data projects are basically IT projects that use a combination of different big data technologies. Since 2012 this topic is attracting researchers, practitioners, and policymakers around the globe. For example, L. Columbus (2016) indicated that the global big data market will grow from $122B in 2015 to more than $187B by 2019. Wikibon’s market forecast report “2016 – 2026 Worldwide Big Data Market Forecast” estimates a compound annual growth rate of 14.4 percent (Wheatley 2016). As an IT project, big data can provide value at several stages: knowledge, organizational agility, business process, and competitive performance (Cörte-Real et al. 2017). According to Gartner, in spite of more than 50 years of history and countless methodologies, advice, and books, IT projects are still failing (Moore 2015). For an enterprise, a big data project is a huge investment that could lead to a crucial change on various levels, especially in terms of the infrastructural part. Thus, introducing a big data project is a strategic decision that needs to be made based on solid understanding of its definition, objectives, risks, and requirements (financial, social, and technical). Often, these issues are highlighted as major challenges when it comes to the realization of big data IT projects. Choosing the right technologies and the right skillset is a major problem (EY 2014; Reid et al. 2015; Wheatley 2016).
Considering these, before the actual project will be realized, enterprises should know whether their project is a big data project or not. Therefore, this research tries to answer the following question: How can the requirements of a project be identified and analyzed to determine the suitability of a combined big data technology application? The main objectives to identify a big data project itself as well as a suitable procedure to determine the most relevant big data requirements and the suitability of the combined use of big data technologies. In order to find a suitable solution to the previously described problem and the formulated research question, the design science methodology according to Peffers et al. (2004) was conducted. This constructional scientific research method is often used to find a possible solution to open organizational problems, such as the success of an IT project (Hevner et al. 2004). Furthermore, to improve the reproducibility and clarity of the investigation, the recommended workflow according to Peffers et al. (2007) was followed which primarily consists of six consecutive steps. The exact sequence is depicted in Figure 1. as well as in the structure of this work. Oriented with a problem-centered intention, the first chapter started with a brief motivation, the problem description, the formulated research question, and its main objectives. In order to find a suitable solution, the theoretical foundations will be determined and discussed in the following second chapter–Related Work. Within the third chapter–Design & Development–all observations and considerations will be combined and applied to find a solution to the main problem. More precisely, a process model will be presented which is based on the Knowledge Discovery in Databases (KDD) process by Fayyad et al. (1996), specific big data requirements, and the approach by Volk et al. (2016). The last chapter contains the demonstration and evaluation of the proposed approach to “observe and measure how well the artifact supports a solution to the problem” (Peffers et al. 2007, p.13). Furthermore, all findings will be discussed and a comprehensive outlook for further research will be given. The paper closes with a short conclusion.

**Figure 1. The recommended design science workflow based on (Peffers et al. 2007)**

**Related Work**

Based on the chosen methodology, the following chapter contains various theoretical considerations which will assist in finding an answer to the previously formulated research question. This is intended to achieve both an overview of the current state of technology as well as to emphasize the importance of a suitable solution. First, Knowledge Discovery in Databases (KDD) by Fayyad et al. (1996) will be described. Using this, a distinction will be made between big data and conventional projects. In the subsequent section, the big data classification framework proposed by Volk et al. (2016) will be presented. The last section contains various theoretical considerations for the requirements engineering process in the context of big data systems.

**Knowledge Discovery in Databases**

KDD essentially describes a multi-staged process which is used to discover knowledge in databases. Starting from the extraction, various preparation methods are used to preprocess, clean, transform, and analyze the data in order to identify certain patterns (Fayyad et al. 1996). These patterns may reveal new insights and knowledge which can later be interpreted and applied, for instance, to other projects and scenarios (Fayyad et al. 1996). The complete process is depicted Figure 2. Although the recommended workflow was published in 1996, it is still applied, independent of the size and structure of the data, as it can be implicitly noticed in (Pääkkönen and Pakkala 2015). For example, today’s projects realized using big data technologies are often implicitly based on KDD. In most of the these application scenarios, however, it is equally indicated that neither the pure implementation nor the choice of the right technologies is a trivial undertaking (Hansmann and Niemeyer 2014; Pääkkönen and Pakkala 2015; Philip Chen and Zhang 2014; Pole and Gera 2016). The constant applicability of the workflow might also be related due to the definition and practical examples of the authors (Fayyad et al. 1996). In addition to numerous areas where analyses for knowledge discovery should be conducted, the authors generalize
individual terminologies and thus emphasize the usage in different areas. According to them “data are a set of facts and pattern is an expression in some language describing a subset of the data or a model applicable to the subset” (Fayyad et al. 1996, p.41) and furthermore that the underlying algorithm itself is predominantly focused on the general analysis, using for example data mining methods. Nevertheless, it is important to note that the process should be conducted only on non-trivial undertakings which can be solved through “a straightforward computation“ (Fayyad et al. 1996, p.41).

![Figure 2. The knowledge discovery in databases process based on Fayyad et al. (1996)](image)

Considering additionally the time dimension, such undertakings can also be viewed as projects which constitute the basic framework for an implementation. According to the Project Management Institute, a project can be described as “a temporary endeavor to create a unique product, service, or result” (Project Management Institute 2008, p.442). Related to methods of IT project management, the main objective of a successful project relies on the combination, implementation, and organization of information technologies (Marchewka 2009). In the context of these observations, one can conclude that big data projects pursue a similar implementation, as other analytical scenarios which will be implemented by using the KDD. However, these differ, above all, in the choice of the necessary technologies and their combination (NIST Big Data Public Working Group 2015). Resulting from this, a big data project can be described as an objective-oriented temporary endeavor with a precisely defined timeframe, whose implementation requires a combined use of specific big data technologies following implicitly the single phases of KDD. However, the decision making at which point in time a big data or a conventional IT project should be realized is similar to the pure identification of the necessary technologies a non-trivial task.

**Classification Framework for Big Data Projects**

Today in most of the definitions of the term “big data”, the individual data characteristics are frequently discussed and described by the V-model (Gandomi and Haider 2015; NIST Big Data Public Working Group 2015). Similar to the definition, the latter differs from contribution to contribution, resulting in ever-new combinations through the extension and modification of existing models (Demchenko et al. 2013; NIST Big Data Public Working Group 2015; Rodríguez-Mazahua et al. 2016) or entirely new conceptions (Bedi et al. 2014). In their core, all of them can be traced back to the three original characteristics of, volume, velocity, and variety, which were already postulated by Doug Laney in (2001). According to his work, volume refers to the amount, variety to the underlying structure, and velocity to the speed of the data which needs to be acquired and analyzed. A first approach which has investigated these characteristics and their severity was described by Volk et al. (2016). In this work, a comprehensive quantitative classification framework was proposed to determine the suitability of a combined big data technology use in IT projects by calculating an assessment value. The complete approach is depicted in Figure 3. In addition to the three core characteristics volume, velocity, and variety which always need to be addressed, the layer-based structure also includes supplementary extensions. Volatility refers to the frequency of structural changes in the data, for instance, if new data sources will be added. The variability, in turn, describes the frequency of changes in the speed of the data flow with which they are acquired and analyzed. The remaining consistency describes the degree of the distribution of the individual systems and the possibility of parallelization, which may be necessary while analyzing the data. In the same way as the three core characteristics, all supporting characteristics were ascertained and tested on their applicability through extensive investigations, using a literature review and use-case analysis (Volk et al. 2016). The framework contains five different levels. While the outer four layers give an exact value dependent of the severity, the fifth level—NULL—is intended for characteristics that cannot be determined presently. If one of the supporting characteristics is assigned to this level, these will not be further utilized.
In the end, all allocated values are summed up and divided by the number of the addressed characteristics which were not assigned to the NULL layer. If the result of the arithmetic mean is greater than or equal to 1.33, an application appears to be meaningful, whereas this is not the case when the value falls below.

Figure 3. A big data classification framework for IT projects based on Volk et al. (2016)

Requirements Engineering in Big Data

In each project, requirement identification is one of the first essential steps. Before starting any projects, the team should understand three fundamentals (Pohl and Rupp 2015): What is requirements engineering? What are the requirements? Who are the project’s stakeholders and what are their roles? In the business context, the IEEE defines requirements as “a condition or capability needed by a user to solve a problem or achieve an objective” (Institute of Electrical and Electronics Engineers 1990, p.62). Every person or organization that impacts the fulfillment of the project’s aim is a stakeholder. These stakeholders, with their different roles, should work together to reach the defined objective. Requirements engineering emphasizes the use of systematic and repeatable techniques that ensure the completeness, consistency, and relevance of the system requirements (Sommerville and Sawyer 1997). It encompasses elicitation, analysis, specification, verification, documentation, and management (Dorfman and Thayer 2000). Due to their novelty, various data characteristics, and complexity, big data projects require new technologies and architectures to enable the stakeholders to create value from the undertaking. This puts forward many challenges, whereby this applies most of all for the identification of the requirements of such a project in the early planning stage. Due to this, it is advisable to achieve constant communication with all stakeholders and shareholders of the potential big data project. Furthermore, a clear distinction between these and conventional projects should be made in various aspects, such as the characteristics of the data and the used technologies. Therefore, differences within the RE process can be expected.

However, after further investigation into this topic, it has been found that only very little research has been carried out so far. This was also particularly highlighted in (Madhavji et al. 2015; Noorwali et al. 2016). Madhavji et al. (2015) emphasize in their work the importance of finding suitable methods with which the crucial data characteristics could be directly addressed, in contrast to classical requirement formulations. A first approach for identifying and formulating big data requirements was proposed by Noorwali et al. (2016). Using a specific notation, it connects traditional quality attributes, such as performance, security, and privacy, with big data characteristics in the following form: “big data characteristic × quality attribute” (Noorwali et al. 2016, p.77). These combinations can be permuted and joined as often as desired with different elements. However, due to the increasing complexity of the compound requirements this conjunction should be restricted to two elements (Noorwali et al. 2016). Additionally, further information about the functional requirements and the possible technological implementation will be needed. According to the authors, the combination of velocity and performance could be described, for instance, through the following requirement: “The system shall use a stream-processing engine with a latency of 0.5 – 2.0 seconds (e.g., Storm, S4, Spark or Samza) to process data in real-time between global earthquake sensors and the data centre” (Noorwali et al. 2016, p.77). As one can easily notice, multiple occurrences of the combined elements are also conceivable and thus other
requirements could be applied to the same combination. For example, the data acquisition may target the same combination, composed of velocity and performance. If no quality requirements can be assigned to a particular combination, this might be a reference to a certain inaccuracy and incompleteness. To prevent later problems, a NULL value will be mapped to the targeted combination. Moreover, the authors recommend in this context a subsequent improvement of this requirement (Noorwali et al. 2016).

**Design and Development**

The following chapter represents the core of the design science methodology. Its main objective represents the creation of an artifact which shall serve as a suitable solution to the initially described problem. Considering the previously explained theoretical principles and the examination of the current state of the art, an extensive model is created within this chapter. Prior to this, the most necessary requirements will be discussed and the strategy with which they can be identified and formulated. The artifact, or rather the solution, is presented in the form of a process model which was composed using Business Process Modeling Notation (BPMN), addressing directly the relevant fundamentals and findings. At the end of this chapter, a high-level framework including the individual components will be presented and described.

**Suitable Requirements**

Considering suitable requirements and the current state of the art, it is important to link functional and quality requirements in order to explicitly take the characteristics of the data and their severity into account (Madhavji et al. 2015; Noorwali et al. 2016). While formulating the relevant requirements, it is advisable to pay attention to the previously described hexagon and its individual layers (see Figure 3). By following the described structure, possible interpretation errors and their corresponding corrections could be prevented during the later assignment. Due to the diversity of the requirements and the characteristics, following the interconnection between functional requirements, data characteristics, and quality attributes, some of which are exemplary described in Table 1. However, it should be noted that not all quality requirements are related to the data characteristics and vice versa. According to Noorwali et al. (2016), in this particular case, certain NULL assignments may exist. The missing or insufficient quality attributes could be due to a shortfall within the requirements engineering process. Hence, similar to the use of the hexagon, a NULL value should be generated and later assigned to the same level, but this applies only to the additional data characteristics. According to the proposed approach by Volk et al. (2016), only the additional characteristics are supplementary. In case of missing quality requirements, especially in the context of the core characteristics, single steps within the requirements engineering process should be retaken. Generally this determination should be as comprehensible as possible, because already one of the six targeted data characteristics has a remarkable influence on the calculation of the decision support (see Figure 3).

<table>
<thead>
<tr>
<th>Char.</th>
<th>Requirements (Characteristic × Compound Req.)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Core Characteristics</strong></td>
<td></td>
</tr>
<tr>
<td>Volume</td>
<td>The system shall store petabytes of data, which need to be processed for positioning of the particles.</td>
</tr>
<tr>
<td>Variety</td>
<td>The system shall analyze the data, which are present in 15 different formats, for fingerprint identification.</td>
</tr>
<tr>
<td>Velocity</td>
<td>The system shall analyze the sensor information in real-time processing speed.</td>
</tr>
<tr>
<td><strong>Additional Characteristics</strong></td>
<td></td>
</tr>
<tr>
<td>Consistency</td>
<td>The distributed system shall adapt to a consistent state from a general weak consistency after a certain event.</td>
</tr>
<tr>
<td>Variability</td>
<td>The system shall scale out during high changes in the data flow, while collecting sensor information.</td>
</tr>
<tr>
<td>Volatility</td>
<td>The system shall react to medium changes in the data format, while identifying new patterns.</td>
</tr>
</tbody>
</table>

**Table 1. Exemplary identification of the necessary requirements**
The Recommended Workflow

In order to be able to define which project requirements can be used to determine the suitability of the application of big data technologies, all theoretical considerations and findings should be brought together. For this purpose, a process model has been developed. This will not solely be used for the assignment of the requirements to the big data classification framework, but involves formal considerations of the description as well. Based on the general planning and realization of such projects, the creation of an initial project idea was chosen as an entry point. In course of this project formulation, the actual origin of the idea does not play a superordinate role, but rather it provides a rough structure for the subsequent requirements analysis. While executing this step, the general project plan and the respective phases of KDD must be considered in addition to the adequate formulation of the requirements. This step of the process will be supported using the previously mentioned scenarios, the modeling of them through use case diagrams, and the notation of specific big data requirements. Additionally, in the case of multiple requirements addressing the same data characteristics, only the requirements with the highest expression should be selected and assigned. As mentioned before, the velocity can be influenced both by the acquisition of the data and its analysis. Thus, possible future adjustments, which might influence a later suitability or certain technologies, are taken into account beforehand. The complete process using BPMN which is created from all previous assertions is depicted in Figure 4.

Figure 4. Identifying initial project considerations (BPMN)

Using this workflow, it is essential to plan the realization of the project as detailed as possible, respecting the course of the various phases in KDD. Hence it will be ensured that the individual requirements are rigorously determined and that subsequent adjustments will not have a negative effect on the further applicability or assessment value of the suitability. This includes the mapping of the highest requirements for the characteristics, the extraction of the layer expressions, and the calculation of the assessment values, according to the approach of Volk et al. (2016). If any of the necessary core characteristics are missing, during the assignment of the specific information, the user will have to revisit the RE process step and determine the missing requirements. Afterwards, the actual calculation can then take place. If a calculated assessment value greater than or equal to the threshold of 1.33 is reached, the combined use of different big data technologies appears to be suitable (see Figure 3). Otherwise, if the assessment values will be lower than 1.33, this may have various reasons. On the one hand, the case may occur that the formulated requirements are not detailed enough, for instance, due to an unspecified scope regarding the processed data. On the other hand, this can also be related to the overall project idea itself. If this fundamental plan is too blurred, it will pervade all of the following steps and will be reflected in the requirements themselves. As a result, the determination of the suitability will be adversely affected. In this case, the user should make further adjustments in the aforementioned steps, in particular the requirements engineering process. However, this can also mean that the realization of the project appears to be feasible through a combination of conventional or an isolated use of big data technologies (Volk et al. 2016). The complete sequence, as previously described, is depicted in Figure 5.
Evaluation

After the main design and development step, the constructed artifact needs to be demonstrated, evaluated, and discussed according to the conducted methodology by Peffers et al. (2007), as depicted in Figure 1. For this purpose, a real-world application scenario was used, derived from a potential project of one of the biggest European B2B trade companies. The main idea of the project is to analyze large quantities of clickstream data in connection with user specific information and their purchase history in order to ensure fast adaptations of the complex category structure and to give user specific product recommendations. After the initial project idea was formulated (Figure 4 - first step), the requirements engineering procedure was carried out (Figure 4 - second step). Using the described framework, all requirements classified as relevant were summarized accordingly and converted into table form, as shown in Table 2.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Layer Value</th>
<th>Compound Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core Char.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Core Char.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Volume</td>
<td>1</td>
<td>The system shall store terabytes of data, consisting of user specific information, clickstream data, and already made purchases.</td>
</tr>
<tr>
<td>Variety</td>
<td>2</td>
<td>The system shall store and analyze the semi-structured data, presented in more than 10 different data formats, for the product recommendation system.</td>
</tr>
<tr>
<td>Velocity</td>
<td>3</td>
<td>The system shall stream the clickstream data directly into the system.</td>
</tr>
<tr>
<td>Additional Char.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consistency</td>
<td>NULL</td>
<td>NULL</td>
</tr>
<tr>
<td>Variability</td>
<td>2</td>
<td>The system shall scale out during irregular changes in the data flow (medium), which will occur during sales, events, and changes in the range of products.</td>
</tr>
<tr>
<td>Volatility</td>
<td>1</td>
<td>The system shall be adaptable and extendable in terms of planned changes in the data structure (low), which will occur after adding new data sources and analyzing methods.</td>
</tr>
</tbody>
</table>

Table 2. Mapping of the requirements to the respective characteristics

More specifically, the first two columns provide information about the type of the affected characteristics. While the third column contains the extracted values of the single layer, the fourth column lists the highest requirement of these specific data characteristics. The values determined from the requirements were visualized for better clarity, similar to the evaluation in (Volk et al. 2016), as shown in Figure 6. After this, the assessment value was calculated. The result of 1.8 revealed that the suitability of a combined big data technology use in this potential project is given. Additionally, the findings of Volk et al. (2016) were applied which state that the individual severity of the data characteristics may provide information about a possible technology recommendation. Since there were no recommendations regarding specific design patterns or technology combinations for the current project, an initial exploration based on these expressions had to be carried out. According to these, an increased technological consideration should occur with regard to the volume and the velocity of the data. Within the scope of this project, the differently structured data needs to be streamed from the sources and analyzed in real-time speed in
In order to be able to react promptly to possible trends, events, and peaks in demands. Due to these requirements, technologies such as Sqoop for data extraction and Hive on Spark for the analysis appear to be promising. Nevertheless, it should be highlighted that partially strong differences between the single phases exist regarding the severity of the data characteristics, such as the described velocity. In future work, it seems to be conceivable to use the original framework in each phase of the big data project or the respective KDD, especially when it comes to the identification of suitable technologies which strongly depends on the uniqueness of single values.

**Figure 6. The illustrated mapping of the targeted layers**

In the course of this work, the application of the described framework from Volk et al. (2016) was solely used once. Only the highest expressions of the data characteristics, according to their respective requirements, were taken into account. By focusing on such individual values and deriving appropriate technologies, various misunderstandings and misapplications may occur. This is the case, for example, when data is extracted at the highest speed level but analyzed periodically. Instead of selecting a technology on the basis of the overall highest expression, the exact phase is more sensible to consider. In the context of technological decisions, the repeated use within the individual phases of KDD to be useful in the future, as depicted in Figure 7. Although interdependencies of the interphase technologies are expected, it is also conceivable that individual recommendations for the specific phases might be discovered. As a result, this would ensure that all big data related requirements will be recognized equally, not just those with the highest expression. This could be realized through the delineation into a matrix and their analysis. Therefore, a reliable technology or a tool could be identified, for example, in the first phase to extract sensor information in real time. Nevertheless, the general suitability of a big data technology combination can still be identified through the determination of the highest values and the application of the hexagon. The achieved assessment value might be even more accurate due to the comprehensive requirements engineering process, compared with the initial approach. In addition to extensive pre-planning, the user also has the opportunity to identify possible design and technology stacks, as analogously described in the context of individual values in (Volk et al. 2016).

**Figure 7. The extended framework allowing possible technology patterns**
In this work, the artifact was evaluated using one use case. Therefore, continues investigations and additional evaluations should be conducted using different methods, such as further use cases’ analysis, thorough technology classifications, and experts’ panel rating. Thus, further developments of the formed requirements as well as the described process appear to be conceivable. As shown in the demonstration and evaluation itself, a suitable allocation option for the hexagon must be ensured. The requirements are not always formulated in such a manner that they can be easily assigned to the respective layer. In the future work, the flexibility should be enhanced by the adaptation of the hexagon itself or by the inclusion of additional steps during the requirements’ extraction phase. Besides the KDD, other approaches to conduct data mining projects exist, such as the Sample, Explore, Modify, Model, Assess (SEMMA) and the CRoss-Industry Standard Process for Data Mining (CRISP-DM) (Azevedo and Santos 2008). In comparison to KDD, the CRISP-DM focuses more on the implementation (Azevedo and Santos 2008). This process should also be investigated and compared to the proposed approach. Beyond the general application of KDD, it was also expected that the requirements will be determined at the very beginning of the process model (cf. Figure 4) or the specific steps (cf. Figure 7). This procedure mainly represents a classical, sequential project management method, such as the waterfall model. Hence, certain incremental adjustments regarding the requirements might be neglected, for instance, if certain quality attributes changed. In the context of the proposed solution, also agile methods must be investigated. By doing this, the complexity of the targeted project should be considered, whereby a general recommendation on which method would be the most suitable will be interesting. In correspondence to this, it is not only recommended to develop such an approach in future works, but also directly compare them to the proposed solution and find a suitable way to determine the right project management method.

Conclusion

Outgoing from the initial problem that more and more projects fail, especially in the context of big data, a suitable solution has been investigated to face major challenges when it comes to the realization of these projects. More precisely, using the design science approach an artifact in the form of a process model was developed, using the compound requirements, a big data classification framework, and Knowledge Discovery in Databases (KDD). Through its use, practitioners should be able to determine and formulate the necessary requirements of big data projects. Based on this information, it is not only feasible to obtain a certain kind of decision support when a choice must be made between conventional and big data technologies in IT projects. An understanding of which requirements are the most relevant will be established as well. Therefore, the awareness of the decisive data characteristics and their influence will be raised implicitly. Furthermore, the proposed process model was demonstrated and evaluated on the basis of a real-world application scenario. Due to its high potential, further development in future investigations appears promising. This is further encouraged through a concluding discussion, during which a comprehensive model is presented to enable both a technology stack recommendation and the general determination of the suitability of a big data use.

References


Project Management Institute 2008. A guide to the project management body of knowledge: (PMBOK guide), Newtown Square, Pa.: PMI.


