Fraunhofer
IESE

FRAUNHOFER INSTITUTE FOR EXPERIMENTAL SOFTWARE ENGINEERING IESE

BIG DATA
MOUNTAINS OF GOLD OR GARBAGE DUMPS?

WHITEPAPER

BIG DATA
MOUNTAINS OF GOLD OR GARBAGE DUMPS?

WHITEPAPER

"Big Data" is on everyone's lips. Many tool vendors and consultants talk about "Big Data" or "Smart Data". But what is behind these buzzwords, and what is the business potential for your organization from making use of Big Data?

The term Big Data economy refers to the vision of data-driven markets of the future. The rapid digitization of products and services (Industry 4.0, Internet of Things) is leading to the creation of large volumes of heterogeneous data in the context of commercial enterprises, civil infrastructures, the Internet, and private households. Successful business models of the future will use these data for creating innovative products and services. In that sense, Big Data is one of the key competencies in future markets – it is indeed the "digital oil of future markets."

With Big Data, everybody hopes for mountains of gold to be gained from their data. Yet, the strategic benefits of Big Data are uncertain, unlike the relatively certain substantial investments that must be made to deploy Big Data, e.g., investments into hardware and software infrastructure and into staff training. In contrast to what is often preached, especially by tool vendors, **there is no universal Big Data solution** and individual solutions available on the market vary widely regarding their capabilities and price. That is why **Big Data solutions should always be adjusted to the particular needs** and abilities of a specific organization. Just like any other initiative that requires organizational change, an honest cost-benefit analysis should precede the deployment of a Big Data initiative [2].

Big Data projects often fail because they are neither aligned with the strategic objectives nor adapted to the operational capabilities of a specific organization. Industrial surveys report that 60% of Big Data projects fail to go beyond piloting and experimentation and are abandoned [3]. Furthermore, less than 50% of lagging organizations have made cultural or business model adjustments that suffice to allow them to benefit from Big Data [1]. Driven by the belief that more data will bring more benefit, organizations focus on creating data lakes without much prior consideration of what they want to achieve with the data and how. Organizations tend to think that they are struggling through a lack of data although in reality most of them have more than enough data to make insightful decisions. The actual issue is that the **data is poorly managed and exploited**. Getting even more data makes the situation only worse. For example, it has been found that 90% of deployed data lakes end up being useless as they are overwhelmed with information assets captured for uncertain use cases [1].

All these problems are even more critical in the context of collaborative enterprises – smart ecosystems – where data-driven business models and Big Data solutions must be aligned with both the individual and the collective interests of multiple organizations. As Bosch concluded in their study [1], data and ecosystems are key drivers of future trends in software engineering and "the challenge isn't the big data [its collection and storage] but the organization's ability to make smart, timely decisions based on the data." It is the ability to collect relevant data, analyze it, and implement data-driven decision making that strongly affects the way a company functions, its architecture, and its workflows.

There is too much data and too few ideas of what to do with it. Turn Big Data into value by aligning it with the business needs and operational capabilities of a particular organization. Business value, not data, is the key driver of Big Data.

The key question today is no longer "whether", but rather how Big Data can help a particular organization achieve their business objectives. Detailed questions that arise in this context include:

- Can your organization benefit from Big Data?
- Which data "treasures" are already available in your organization and how have they been exploited so far?
- Are there any hidden business potentials in the available data that have not been exploited so far?
- Are the required competencies available?
- Is the quality of the data sufficient to realize the anticipated business benefits?
- What investments are necessary for deploying a Big Data initiative?
- How does Big Data affect existing business models?

A thorough analysis of the Big Data business potential provides answers to these questions, ideally before any potentially wrong investments have been made to introduce Big Data in an organization.

Analysis of Big Data Potential: How to demonstrate the business value of Big Data?

As in the case of any other initiative that requires organizational change, the key question "Why?" should be answered before the introduction of a Big Data initiative. Following the thoughts of Michael E. Porter from Harvard Business School, who pointed out that "Success requires both the right strategy and operational effectiveness", it first needs to be made clear what is going to be achieved with Big Data and how. In other words, a clear vision of the desired business impact must shape the integrated approach to data, technology, competencies, and organizational transformation. Before deploying any specific solution, one must be able to demonstrate the link between data and its potential business value, and determine any gaps in the operational capabilities (technologies, infrastructure, competencies, etc.) that are required for realizing the business value. In the end, the overall business potential of Big Data is the difference between the expected business benefit and the cost of realizing an appropriate Big Data solution.

Success requires both the right strategy and operational effectiveness.

[Michael E. Porter, Harvard Business School]

Identify and close the gap between the potential benefits of available data and the ability to turn that data into business insight and value.

In practice, the ultimate goal of employing Big Data is business innovation. We aim at using Big Data to improve an existing business model or to create a new business model. The success of a Big-Data-driven business model is measured through the value it delivers and the cost it requires to be realized, including the implementation of a specific Big Data solution. To be successful, Big Data must create significantly more value than costs.

In general, Big Data can create business value in any business model area, including customer relationships, product excellence, and operative excellence (see Figure 1). For instance, Big Data can provide support for these activities:

- Identifying and addressing new customers or unknown needs of current customer segments;
- creating new attractive products and services;
- accelerating internal processes for creating and delivering value, and for capturing value from customers.

Figure 1 Business value creation areas of a business model

Data-driven innovation: How to innovate a business model with Big Data?

As of today, two main types of strategy are being followed to exploit data for creating business value (benefit), that is, to innovate business using data: data-driven and business-driven strategies (see Figure 2). Although their starting points are very different, both aim at finding innovative ideas for filling a business need; that is, for solving business problems and/or exploiting business opportunities.

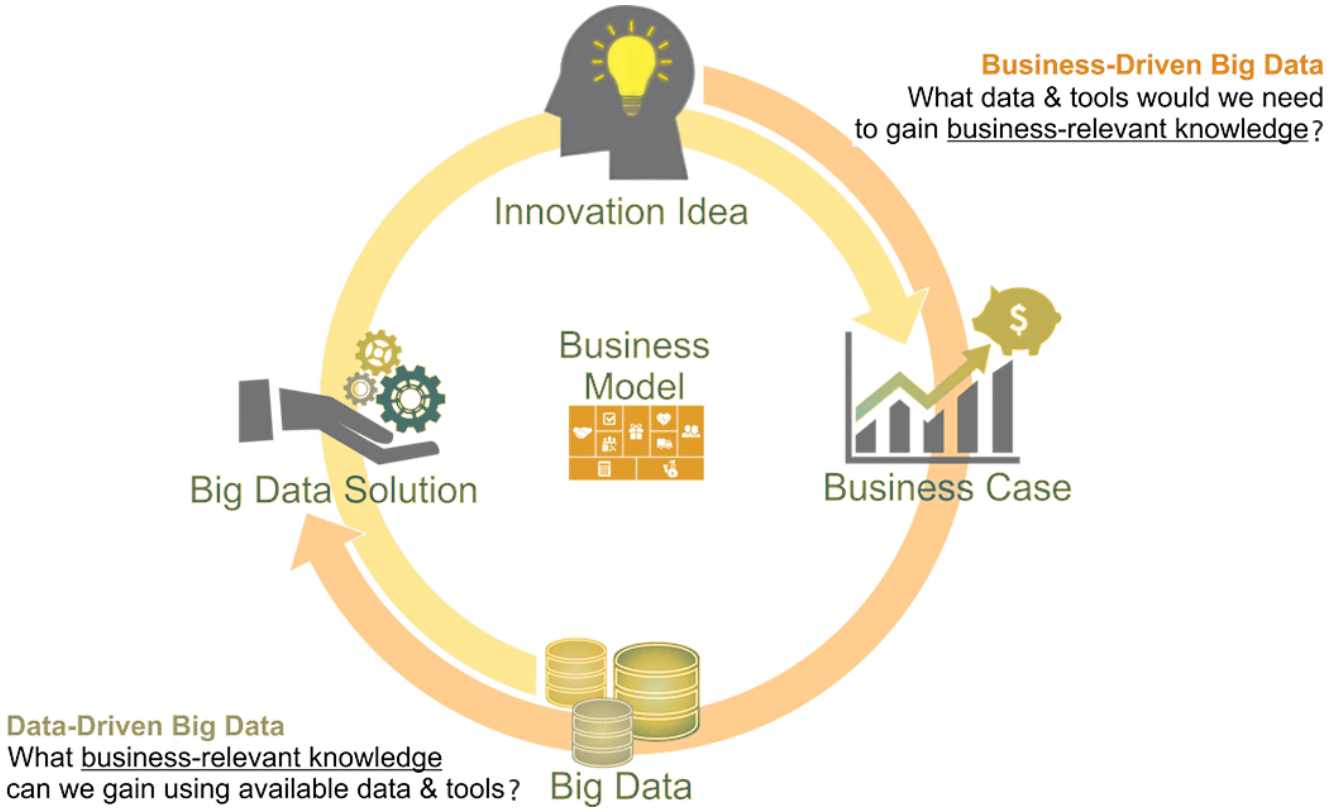

Figure 2 Business- and data-driven Big Data strategies

The **data-driven strategy** takes whatever data is available and explores it using Big Data solutions built up in an ad-hoc manner to gain business-relevant knowledge. The **business-driven strategy** starts with the business needs and derives potential business cases in order to come up with appropriate Big Data solutions for gaining relevant business knowledge. Both strategies have their strengths and weaknesses (see Table 1) and thus should be considered as complementary rather than competitive in practice.

Business-driven Big Data	Data-driven Big Data
(+) Goal-oriented: The Big Data solution is driven by business goals and needs and fits the capabilities and constraints of a specific organization.	**(+) Data-driven:** The Big Data solution is driven by the data to be analyzed and, to a large extent, must fit the available capabilities and constraints. In the case of a large capability gap, necessary technological and human resources can be acquired for a limited time of explorative analysis from outside the organization (e.g., cloud services and external consultants).
(–) Tunnel vision: Big Data solution ideas are likely to be limited due to known business needs and context. Although business needs and potential solutions are identified with the help of creativity techniques (out-of-the-box thinking), relevant ideas (e.g., business opportunities) may be missed.	**(+) Open mind:** Big Data solutions highly depend on available data. Explorative data analysis has the advantage that it may lead to disruptive insights that could otherwise never be gained from domain experts (e.g., by using creativity techniques).
(+) Systematic: The Big Data solution is based on well-founded procedures and a traceable selection of alternative options and implementation decisions.	**(–) Unsystematic:** Explorative data analysis typically has an unsystematic character. Data scientists explore data using familiar techniques in a trial and error process, guided by their own analysis experiences and available resources (financial and technological).
(+) Foreseeable cost-benefit ratio: The potential benefits and costs of implementing a specific Big Data solution are assessed prior to actually deploying the solution in the target environment.	**(–) Unforeseeable cost-benefit ratio:** The benefits of the data-driven Big Data approach can only be estimated in retrospect, after data exploration has been completed and the gained insights (incl. their potential business value to the organization) are known.
(+) Foreseeable risk: The Big Data solution is aligned with the business needs and operational capabilities of a specific organization, and evaluated gradually prior to being deployed in the productive environment.	**(–) High risk:** The business value of Big Data solutions (incl. data) can hardly be forecasted upfront and can only be estimated in retrospect. The value can range from disruptive insights to no relevant insights at all.
(–) Effort-intensive: Although foreseeable, the cost of business-driven Big Data can be significant. Deriving and assessing relevant Big Data business cases within the scope of a Potential Analysis and stepwise deployment of a Big Data solution requires appropriate time and resources.	**(+) Minimal effort:** The Big Data solution is planned ad-hoc according to a fixed budget and assigned to explorative analysis.

Table 1 Strengths and weaknesses of business-driven and data-driven Big Data strategies

One way of combining both strategies in practice is to involve both domain expertise and the results of a data analysis for the derivation of Big Data business cases. For example, representatives of business experts and analysts can search for innovative ideas during a joint creativity workshop. The business experts can contribute their knowledge and understanding of business, whereas the analysts can share their knowledge of available data and existing analysis results. Innovation ideas can then be concretized in the form of **business cases** that specify innovative **business solutions** based on data.

Big Data Business Case

A Big Data business case initially, yet clearly, shows the path towards an innovative business model based upon Big Data and provides arguments for implementing this path by comparing it to alternative paths (so-called solution options) that are not based on Big Data. Alternative business innovation solutions are compared based upon detailed information, including:

- Assumptions upon which the solution's ability to achieve the business outcome is based
- Benefits expected after implementing the solution
- Costs required to implement the solution
- Time for implementing the solution and gaining benefits
- Risks associated with the solution that may impact the realization of the business benefit

In order to get a chance of being deployed, the Big Data solution must demonstrate a cost-benefit ratio that is significantly better than that of alternative solutions.

Based upon the business case, appropriate **Big Data solutions** – including the necessary data, technologies, infrastructure, and competencies – are derived, taking into account organizational capabilities and constraints such as available financial and human resources or existing infrastructure and competencies. Data-driven business innovation is an iterative process (see Figure 3) in which the business innovation ideas and the associated Big Data solutions are evaluated and matured. In each iteration, viable prototypes of the business innovation ideas and Big Data solutions are designed and tested against success criteria such as business profitability or technical feasibility. Based on the test feedback, business innovation ideas and the associated Big Data solutions are improved. Since the main objective of these improvement cycles is to learn about the underlying problem and explore alternative solutions, one should not be afraid to discard unpromising ideas.

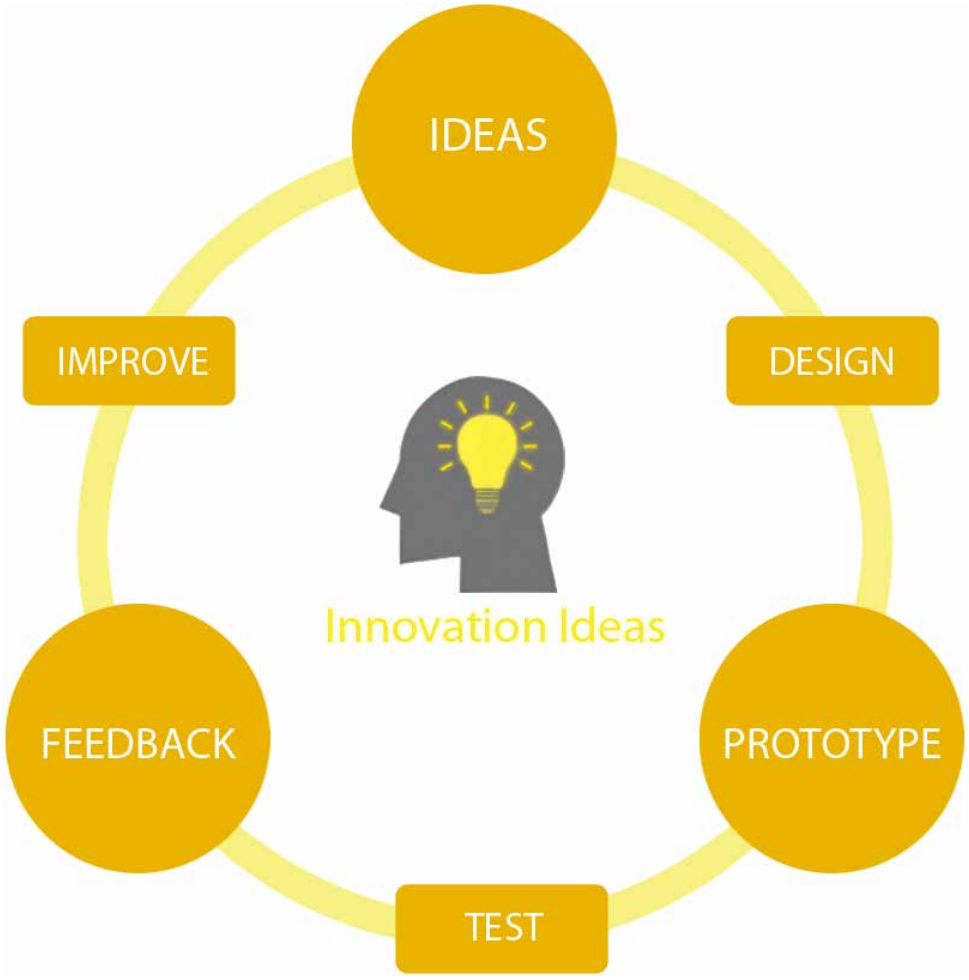

Figure 3 Business innovation process based upon the Lean Startup strategy [4].

Analysis of Big Data Business Potential: How to identify and close the gap between the potential benefits of data and the ability to turn that data into business value?

In each iteration of the business innovation cycle, the business potential of a specific Big Data solution is assessed in the light of the business value it is expected to create and the cost required to realize it. In that sense, the analysis of Big Data business potentials ("**Potential Analysis**" for short) supports the derivation, evaluation, and maturation of business innovations based on Big Data in a way that helps to minimize the risk and potential loss caused by investing in Big Data solutions that do not create the expected business value.

The core element of the potential analysis is a concrete big data **business case** (see Figure 4). It specifies a particular Big-Data-driven business innovation, including its context with the underlying business need, the **business solution** with the expected benefits (value), and the **Big Data solution** with the organizational **readiness** needed to implement it. The outcomes of the potential analysis answer key questions that should precede any Big-Data-driven change: *In what context should it happen? What business benefit (value) should be gained with Big Data? Are the organizational capabilities needed for implementing an appropriate Big Data solution available? How much will the deployment of an appropriate Big Data solution cost?*

Figure 4 Analysis of Big Data business potentials

A potential analysis starts with the specification of the organizational **scope and current situation**. Its purpose is to understand the organization's internal and external factors that (1) are the source of potential business challenges and opportunities and (2) influence the feasibility of potential Big Data solutions.

The organization's internal and external situation, including cits urrent business model, creates the basis for generating business innovation ideas. In essence, we are looking for potential data-driven **business solutions** to improve existing or create new business models. Each business innovation idea (potential business solution) is subject to an **evaluation** regarding its chances of success and the organizational **readiness** needed to realize it. Example questions to answer at this point include: *Is the considered business challenge or opportunity associated with relevant customer needs? Are data-driven products and services that address the need desired by the intended customers? Will a data-driven business model generate the expected profits?* Based on the outcomes of the evaluation, potential business solutions are improved or abandoned. The **business solutions** with the highest chances of success are summarized in the form of data-driven business cases. A **business case** not only documents a specific business need and the proposed data-driven business solution but also motivates its further realization in the form of a Big Data solution by comparing it to alternative business solutions, e.g., solutions not based on Big Data. Only after a specific business solution has shown a sufficiently high chance of success does the innovation process continue with the development of a corresponding **Big Data solution**, which is the technical realization of the business solution idea.

Big Data Solution Concept

A Big Data solution concept specifies a Big Data usage scenario and its technical implementation, in particular:

- the target usage scenario for Big Data as defined in the associated business case,
- the target scenario stakeholders and their information needs,
- the solution idea of how to use the data to address the specified information needs and assumptions underlying the idea
- the required data and its quality requirements,
- the data preparation and analysis approach,
- the criteria for evaluating the quality of the data analysis outcomes
- the Big Data architecture and technologies.

The motivation for developing a business solution first is that it is typically significantly cheaper to test and (potentially) fail with it than with a Big Data solution, where prototyping and testing already require considerable investments into infrastructure and staff. The development of a Big Data solution includes deciding about the Big Data methods, the infrastructure, and the skills required to realize the business solution. Similar to the business solution, the Big Data solution evolves through test-feedback-improve evaluation cycles. In each **evaluation** cycle, the **readiness** of the organization to implement and deploy a specific Big

Data solution is assessed and the gap between the required and the available organizational capabilities is determined. Example assessment questions are: *Are the required data available and do they have sufficient quality? Can the intended analysis methods be applied? Is the required hardware and software infrastructure available? Are human resources available that have the necessary competencies?* A potential capability gap can be addressed by either adjusting the business solution and/or the associated Big Data solution (i.e., necessary data, analysis methods, and infrastructure).

To minimize the risk of failing with Big Data and to reduce potential business losses, the Big Data solution concept is evaluated and matured throughout several specific lab and piloting stages before it can be deployed in a productive environment (see Figure 5). After each step, the results are evaluated according to pre-defined criteria. Note that although the evaluation of the Big Data solution focuses on the technical aspects, the business aspects are still considered as well. For instance, the accuracy of a prediction algorithm may directly influence the reliability of customer-relevant forecasts and thus affect the desirability of the realized business solution for the customer.

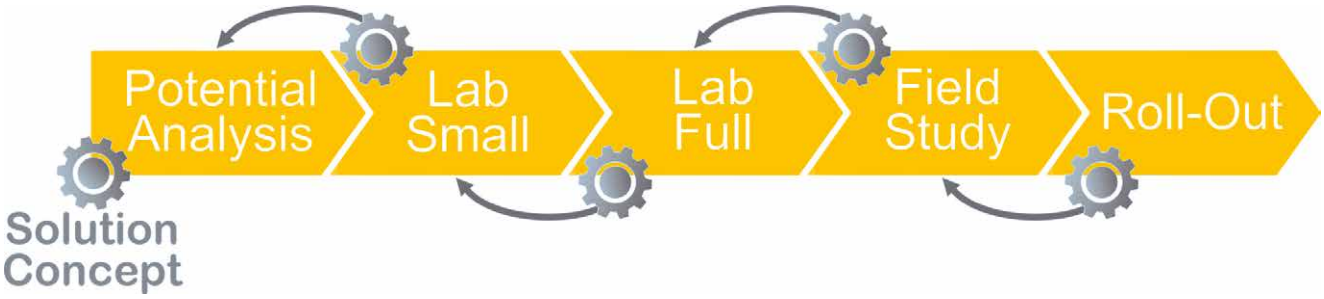

Figure 5 Staged realization of Big Data approach

At each stage, the Big Data solution concept, the associated business case, and the organizational readiness are revised based upon the evaluation outcomes. In the very first stage, an initial Big Data solution concept is blueprinted and evaluated conceptually without any practical implementation. In the following stages, specific "in-use" aspects of the Big Data solution concept are verified after being implemented in test environments. For example, the performance and the scalability of the selected data analysis technologies are evaluated in a lab environment using real or simulated data. Evaluating integration with existing infrastructure and processes as well as user acceptance, on the other hand, requires piloting the Big Data approach in the intended target environment, yet within a limited scope in order not to threaten the operation of the productive system in case anything goes wrong. Based on the outcomes of the implementation, the solution concept and the corresponding business solution are revised and re-assessed regarding their business impact and organizational readiness. Only if the Big Data solution has successfully run through all intermediate stages is it rolled out into a productive environment. Yet this is not the end; as the organizational context changes continuously, the effectiveness of the realized data-driven business model should be revised on a regular basis.

Bottom Line: Start a Big Data journey by asking "Why?"

Summarizing, the **potential analysis** aims at finding a Big Data solution for a specific business innovation idea that provides the best return on investment, that is, a solution that promises to deliver the best trade-off between the potential business benefits and the investments required for deploying it. The development of a suitable solution may require several iterations in which both the anticipated data-driven business solution and the required Big Data solution concept are revised.

As shown by various recent studies, basing business decisions on data makes much more commercial sense than relying on gut feelings and opinions. Data-driven organizations are typically more successful than competitors that do not rely on data. Yet, like many other technology-dependent changes, the failure rate for those on this journey is considerable. Organizations should spend more time asking themselves **why** they need Big Data rather than rushing into technological solutions they believe to be silver bullets. This simple question should trigger detailed follow-up questions, all of which need to be considered thoroughly in the context of current business needs and constraints. Only after we know what we want to achieve with Big Data should we continue by asking **how** we can achieve it. The "how" question should then lead us to the required data and technologies. The analysis of Big Data business potentials we sketched above provides guidance to organizations by answering the "why" and "how" questions regarding Big Data.

Our Services

- **Analysis of Big Data business potentials:** We provide you with guidance by analyzing the potential business value of Big Data for your organization and by selecting the right Big Data approach.
- **Business-driven analysis of available data:** We analyze your data for you from the perspective of your business questions to check how much business potential your current data has.

Our Seminars

- **Big Data business potentials:** This two-day seminar offered by Fraunhofer IESE introduces Big Data with its opportunities and threats, and shows how to analyze the business potentials of Big Data in the context of a specific organization. Theoretical material presented during the seminar is illustrated with the help of an intuitive real-world example. On the second day, the participants have an opportunity to apply the learned material in practice during a joint potential analysis workshop.
  https://www.iese.fraunhofer.de/de/seminare_training/analyse_des_potenzials_von_big_data.html

- **Data Scientist Basic:** Five-day intensive course with the possibility of getting the "Data Scientist Basic" certificate. This seminar is offered by the Fraunhofer Big Data Alliance (including Fraunhofer IESE) and provides a comprehensive overview of relevant Big Data topics such as Big Data architectures, data preparation and analysis, data management, data security and privacy, data visualization, and analysis of Big Data business potentials.
  http://www.bigdata.fraunhofer.de/de/datascientist/seminare/zertifizierung.html

- **Data Analytics – Importance, Potentials, Realization:** Two-day seminar offered jointly by Fraunhofer IESE and Fraunhofer IAIS. The objective of this seminar is to provide business and project managers with inspiration on how to gain business benefits from Big Data. For this purpose, the seminar provides a practice-oriented introduction to relevant Big Data concepts and illustrates them with a number of real-world application examples.
  https://www.iais.fraunhofer.de/de/geschaeftsfelder/big-data-analytics/uebersicht/data-scientist-schulungen/data-analytics-potentials.html

BIG DATA
MOUNTAINS OF GOLD OR GARBAGE DUMPS?

WHITEPAPER

Bibliography

1. J. Bosch, "Speed, Data, and Ecosystems. The Future of Software Engineering." IEEE Software, vol. 33, no. 1, January/February 2016, pp. 82–88.

2. J. Heidrich, A. Trendowicz, C. Ebert: Exploiting Big Data's Benefits. IEEE Software, vol. 33, no 1. July/August 2016, pp. 111-116.

3. N. Heudecker et al., "Predicts 2015: Big Data Challenges Move from Technology to the Organization", Gartner Report, November 2014.

4. E. Ries, The Lean Startup: *How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses.* Crown Business; 1st edition, September 2011.

Contact

Dr. Adam Trendowicz
adam.trendowicz@iese.fraunhofer.de
Telefon +49 631 6800-2137
www.iese.fraunhofer.de

Fraunhofer Institute for Experimental Software Engineering IESE

Fraunhofer-Platz 1
67663 Kaiserslautern

Directors

Prof. Dr.-Ing.
Peter Liggesmeyer
Prof. Dr. Dr. h. c.
Dieter Rombach

Fraunhofer Institute for Experimental Software Engineering IESE

About Fraunhofer IESE

For more than 20 years, Fraunhofer IESE in Kaiserslautern has been one of the worldwide leading research institutes in the area of software and systems engineering methods. It addresses innovation topics such as Industry 4.0, Big Data, and Cyber-Security. The institute is a technology and innovation partner for digital transformation in the areas of Autonomous & Cyber-Physical Systems and Digital Services and performs research on the interaction between embedded systems and information systems in digital ecosystems.

Short Bio Adam Trendowicz

Adam Trendowicz is a senior consultant in data engineering at the Fraunhofer Institute for Experimental Software Engineering IESE in Kaiserslautern, Germany. At Fraunhofer IESE, Dr. Trendowicz is responsible for the Big Data competence area with a focus on business-value-oriented analysis of Big Data potentials. He received his master's degree in software engineering from Poznan University of Technology (Poland) and his PhD in computer science with a thesis on software prediction models from the University of Kaiserslautern (Germany).

2017-09