Transforming manufacturing companies towards a data-driven enterprise – A resource-based view perspective

SVEN KOPERA
Transforming manufacturing companies
towards a data-driven enterprise – A
resource-based view perspective

by

Sven Kopera
Att transformera tillverkningsföretag till ett datadrivet företag - Ett resursbaserat perspektiv

Sven Kopera
Abstract

Background: Big Data is the new lever of gaining competitive advantage for manufacturing companies. However, several challenges pose barriers to a successful implementation. Especially, the integration, collection and analysis of data originating from a wide variety of data sources is one of the main challenges but at the same time of high importance for effective and insightful Big Data analytics.

Purpose: This thesis’ aim was to investigate which resources manufacturing companies have to acquire and which capabilities have to be developed to overcome current challenges related to the beneficial utilization of Big Data analytics using internal and external data sources.

Theoretical framework: Drawing on the resource-based view and its related concepts of resources, capabilities and dynamic capabilities, a conceptual research framework was developed.

Methods: A qualitative survey was conducted using semi-structured interviews. The sample included representatives from Big Data technology suppliers, manufacturing companies and industry experts from a management consultancy. Questions related to data discovery, data integration and data analysis were asked as well as challenges to be overcome during implementation.

Results: Overall, 176 potential interview partners were contacted out of which 27 agreed to an interview, yielding a response rate of 15 per cent. The interviews revealed that manufacturing companies consider both new internal and external data sources for their Big Data analytics. Moreover, the interviewees identified 21 challenges as well as possible resources and capabilities to tackle them on four levels, the managerial, technological, organizational and individual level respectively.

Discussion: To successfully reach data integration, manufacturing companies must pass through four stages of data archetypes. Transition between archetypes is enabled by developing dynamic capabilities since resources and capabilities in each archetype are different.

Conclusion: The findings contribute to the developing body of knowledge regarding Big Data research from a management perspective by highlighting necessary capabilities and resources to overcome current challenges.

Key-words
Big Data, Manufacturing, Big Data challenges, Big Data capabilities, Big Data archetype
Sammanfattning

**Bakgrund:** Big data är den nya tillgången för att få konkurrensfördelar för tillverkningsföretag. Men flera utmaningar utgör hinder för en framgångsrik implementation. Integration, insamling och analys av data från en mängd olika datakällor är de främsta utmaningarna och samtidigt av stor betydelse för effektiv och insiktsfull Big Data-analyse.

**Syfte:** Uppsatsens syfte är att undersöka vilka resurser tillverkningsföretag måste införskaffa och vilka förmågor som måste utvecklas för att överkomma de nuvarande utmaningarna med användandet av Big Data-analyse med hjälp av interna och externa datakällor.

**Teoretiskt ramverk:** Med utgångspunkt i det resursbaserade perspektivet och dess relevanta begrepp resurser, förmågor och dynamiska förmågor utvecklades en konceptuell forskningsram.

**Metoder:** En kvalitativ undersökning genomfördes med hjälp av semi-strukturerade intervjuer. I urvalet ingår representanter från Big Data-teknikleverantörer, tillverkningsföretag och branschexpeter från en managementkonsultfirma. Frågor relaterade till datainsamling, dataintegration och dataanalys ställdes samt frågor om utmaningar att överkomma under implementeringen.

**Resultat:** Sammanlagt kontaktades 176 potentiella intervjuobjekt, varav 27 gick med på en intervju, vilket gav en svarsfrekvens om 15 procent. Intervjuerna visade på att tillverkningsföretag överväger både nya interna och externa datakällor för sina Big Data-analys. Vidare identifierade intervjuerna 21 utmaningar samt möjliga resurser och förmågor att hantera dem på fyra nivåer; på ledningsnivå, teknisk, organisatorisk och individnivå.

**Diskussion:** För att framgångsrikt kunna nå dataintegration måste tillverkningsföretag passera genom fyra stadien av dataarketyper. Övergången mellan arketyper möjliggörs genom att utveckla dynamiska förmågor eftersom resurser och förmågor för varje arketyp är olika.

**Slutsats:** Resultaten bidrar till att utveckla kunskapen om Big Data forskning ur ett ledningsperspektiv genom att framhäva nödvändiga resurser och förmågor för att klara de nuvarande utmaningarna.

**Nyckelord**
Big Data, Manufacturing, Big Data challenges, Big Data capabilities, Big Data archetype
1 Introduction ............................................................................................................. 1
  1.1 Background ........................................................................................................ 1
  1.2 Problematization ................................................................................................. 2
  1.3 Purpose ............................................................................................................... 3
  1.4 Research questions ............................................................................................ 4
  1.5 Research implications ......................................................................................... 4
    1.5.1 Rationale for research .................................................................................. 4
    1.5.2 Originality and expected contribution ......................................................... 5
    1.5.3 Scope of research ......................................................................................... 6
  1.6 Disposition of thesis ......................................................................................... 6

2 Introduction to Big Data in manufacturing ......................................................... 8
  2.1 Enabling technologies for Big Data in manufacturing ....................................... 8
    2.1.1 Industry 4.0 .................................................................................................. 8
    2.1.2 Internet of Things ......................................................................................... 9
    2.1.3 Cyber physical systems ................................................................................ 10
    2.1.4 IT systems and Big Data platforms ............................................................. 10
    2.1.5 Manufacturing and Big Data ...................................................................... 12
  2.2 Characteristics of Big Data .............................................................................. 13
    2.2.1 The five Vs of Big Data .............................................................................. 13
    2.2.2 Big Data value chain .................................................................................. 14
    2.2.3 Data discovery ............................................................................................. 14
    2.2.4 Data integration ........................................................................................... 17
    2.2.5 Data exploitation ......................................................................................... 18

3 Theoretical framework ...................................................................................... 20
  3.1 Resource-based view of the firm ...................................................................... 20
    3.1.1 Resources, capabilities and competencies .................................................. 20
    3.1.2 Dynamic capabilities .................................................................................. 21
    3.1.3 Competitive advantage .............................................................................. 23
  3.2 Mapping the theory to Big Data ...................................................................... 24
  3.3 Big Data challenges, resources and capabilities ............................................ 24
  3.4 Conceptual research framework ..................................................................... 25

4 Methods ........................................................................................................... 27
  4.1 Primary and secondary sources ...................................................................... 27
  4.2 Scoping review ................................................................................................. 27
  4.3 Qualitative survey ............................................................................................ 27
    4.3.1 Research design & purpose ..................................................................... 27
    4.3.2 Logic .......................................................................................................... 28
    4.3.3 Process ....................................................................................................... 28
    4.3.4 Outcome .................................................................................................... 31
  4.4 Overview of interviews ................................................................................... 32
    4.4.1 Response rates and dropout analysis ......................................................... 32
    4.4.2 Technology suppliers ............................................................................... 32
    4.4.3 Manufacturing companies ....................................................................... 34
    4.4.4 Experts ....................................................................................................... 34
  4.5 Critical evaluation ............................................................................................ 34
    4.5.1 Internal validity .......................................................................................... 34
    4.5.2 Construct validity ....................................................................................... 35
    4.5.3 External validity ........................................................................................ 35
    4.5.4 Reliability .................................................................................................. 35
5 Results and analysis ................................................................. 36
5.1 Structure of results ............................................................... 36
5.2 Data sources for Big Data ....................................................... 36
  5.2.1 Data sources and types ....................................................... 36
  5.2.2 Purpose of acquiring new data sources .................................. 38
  5.2.3 Analysis ................................................................. 40
5.3 Data discovery ....................................................................... 41
  5.3.1 Collection of data sources .................................................... 41
  5.3.4 Data access and preparation ................................................ 44
5.4 Data integration .................................................................... 48
5.5 Data exploitation .................................................................... 50
5.6 Summary of results ............................................................... 53
5.7 Analysis of identified challenges, resources and capabilities ......... 54
6 Discussion ................................................................................. 62
  6.1 Research findings in relation to data sources ............................. 62
  6.2 Research findings in relation to Big Data challenges ................... 63
  6.3 Research findings in relation to resources and capabilities ........... 68
    6.3.1 Static view ............................................................... 68
    6.3.2 Dynamic view ............................................................ 68
  6.5 Reflections on sustainability .................................................... 72
7 Conclusion and recommendations ................................................ 74
  7.1 Reviewing the purpose and answering research questions .......... 74
  7.2 Research contribution ........................................................... 75
    7.3.1 Theoretical contribution .................................................. 76
    7.3.2 Empirical and practical contribution .................................... 76
  7.4 Limitations and further research .............................................. 78
8 Bibliography .............................................................................. I
9 Appendix ...................................................................................... A
Table of Figures

FIGURE 1: OVERVIEW OF INDUSTRY 4.0 AND ITS TECHNOLOGICAL CONCEPTS (SOURCE: ROLAND BERGER) ...9
FIGURE 2: VOLUME, VELOCITY AND VARIETY OF BIG DATA BY RUSSOM (2011), OWN VISUALIZATION ..........13
FIGURE 3: DATA VALUE CHAIN FRAMEWORK BY MILLER & MORK (2013), OWN VISUALIZATION ..............14
FIGURE 4: DATA SOURCES AS IDENTIFIED BY HARTMANN ET AL. (2016), OWN VISUALIZATION ..............16
FIGURE 5: CONCEPTUAL RESEARCH FRAMEWORK ..............................................................................26
FIGURE 6: VISUALIZATION OF ALL IDENTIFIED CHALLENGES ACROSS FOUR DIMENSIONS ALONG THE VALUE CHAIN ........................................................................................................26
FIGURE 7: BIG DATA ARCHETYPE MATRIX .......................................................................................69
Table of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1:</td>
<td>BIG DATA CHALLENGES</td>
<td>25</td>
</tr>
<tr>
<td>Table 2:</td>
<td>BIG DATA RESOURCES AND CAPABILITIES</td>
<td>25</td>
</tr>
<tr>
<td>Table 3:</td>
<td>IDENTIFIED THEMES AND CODES OF DATA ANALYSIS</td>
<td>31</td>
</tr>
<tr>
<td>Table 4:</td>
<td>DESCRIPTIVE STATISTICS OF INTERVIEW RESPONSE RATES FOR EACH SAMPLE GROUP</td>
<td>32</td>
</tr>
<tr>
<td>Table 5:</td>
<td>DESCRIPTIVE STATISTICS OF INTERVIEW RESPONSE RATES FROM TECHNOLOGY SUPPLIERS</td>
<td>32</td>
</tr>
<tr>
<td>Table 6:</td>
<td>INTERVIEWED PERSONS AT TECHNOLOGY SUPPLIERS</td>
<td>33</td>
</tr>
<tr>
<td>Table 7:</td>
<td>INTERVIEWED PERSONS AT MANUFACTURING COMPANIES</td>
<td>34</td>
</tr>
<tr>
<td>Table 8:</td>
<td>INTERVIEWED PERSONS AT A MANAGEMENT CONSULTANCY</td>
<td>34</td>
</tr>
<tr>
<td>Abbreviation (Abb)</td>
<td>Full Form</td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>BDA</td>
<td>Big Data analytics</td>
<td></td>
</tr>
<tr>
<td>BI</td>
<td>Business Intelligence</td>
<td></td>
</tr>
<tr>
<td>BI&amp;A</td>
<td>Business intelligence and analytics</td>
<td></td>
</tr>
<tr>
<td>B2B</td>
<td>Business to Business</td>
<td></td>
</tr>
<tr>
<td>B2C</td>
<td>Business to customer</td>
<td></td>
</tr>
<tr>
<td>CPS</td>
<td>Cyber physical system</td>
<td></td>
</tr>
<tr>
<td>CRM</td>
<td>Customer relationship management</td>
<td></td>
</tr>
<tr>
<td>DC</td>
<td>Dynamic capabilities</td>
<td></td>
</tr>
<tr>
<td>DDD</td>
<td>Data-driven decision-making</td>
<td></td>
</tr>
<tr>
<td>ERP</td>
<td>Enterprise resource planning</td>
<td></td>
</tr>
<tr>
<td>IIoT</td>
<td>Industrial Internet of Things</td>
<td></td>
</tr>
<tr>
<td>IoT</td>
<td>Internet of Things</td>
<td></td>
</tr>
<tr>
<td>IP</td>
<td>Internet Protocol</td>
<td></td>
</tr>
<tr>
<td>IS</td>
<td>Information System</td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td>Information Technology</td>
<td></td>
</tr>
<tr>
<td>LAN</td>
<td>Local Area Network</td>
<td></td>
</tr>
<tr>
<td>MDMS</td>
<td>Master data management system</td>
<td></td>
</tr>
<tr>
<td>MES</td>
<td>Manufacturing execution system</td>
<td></td>
</tr>
<tr>
<td>MRQ</td>
<td>Main research question</td>
<td></td>
</tr>
<tr>
<td>M2M</td>
<td>Machine-to-machine</td>
<td></td>
</tr>
<tr>
<td>OT</td>
<td>Operations Technology</td>
<td></td>
</tr>
<tr>
<td>RBV</td>
<td>Resource based view</td>
<td></td>
</tr>
<tr>
<td>RFID</td>
<td>Radio Frequency Identification</td>
<td></td>
</tr>
<tr>
<td>RQ</td>
<td>Research question</td>
<td></td>
</tr>
</tbody>
</table>
Foreword

Reflecting on the last seven months writing this thesis, I realized that carrying out the thesis project by myself has had both its advantages and disadvantages. Being only by myself, I quickly realized that a partner to discuss ideas, obstacles and issues with as well as argue about certain aspects would have been helpful from time to time. At the same time, being solely responsible for this project, I did not only learn a lot about the topic of Big Data, but improved my organizing, project management and time management skills. I am grateful for this learning journey, from which I will benefit in my future working life.

- Sven Kopera
Stockholm, September 2017

Acknowledgements

I would like to thank the following persons for supporting me with their advice, guidance and feedback throughout this thesis project:

- My seminar leader, Mats Engwall, for challenging my views and thoughts on the projects and pushing me to apply different perspectives
- My supervisor from KTH, Jannis Angelis, for his support, sharing his research knowledge and feedback on structural improvements
- My TINEM program director of studies, Lars Uppvall, for providing the necessary foundations in research methods to carry out this thesis project
- My seminar group for great discussions and feedback sessions
- My interviewees who have willingly taken time to share their experiences and thoughts with me during the interviews

Special thanks to my commissioning company and my external supervisor for supporting me throughout this thesis project. I hope my research findings can provide some new perspectives and views on the topic of Big Data in manufacturing companies.
1 Introduction

1.1 Background

In this introductory chapter, the background of the research topic is briefly explained from which the problematization is derived. Subsequently, the research purpose and questions, are addressed. At last, the expected contribution, delimitations and limitations of the research are discussed. The whole thesis is outlined in the disposition.

1.1 Background

This study addressed how challenges associated with the integration of multiple data sources for Big Data analyses in the manufacturing industry can be overcome by acquiring resources and developing capabilities.

Today, in a time of the so-called digital era, more and more digital data are available to be accessed by individuals and the trend of data generation is likely to increase in the future (Chong & Shi 2015; Addo-Tenkorang & Petri T Helo 2016). The amount of digital data has been growing rapidly in the last two decades (Chen et al. 2014). It is predicted that the ‘digital universe’ will double in size every two years resulting in up to 44 zeta (10\(^{21}\)) bytes of digital data by the end of 2020 (IDC 2014).

Technological improvements of digital sensors, communication devices, computation power and data storage have contributed to the data increase (Tien 2013). Nowadays, data originate from a steadily increasing variety of sources (Hartmann et al. 2016). The large amount of data generated from many different sources and in many different types refers to the concept of Big Data: “With this overwhelming amount of complex and heterogeneous data pouring from any-where, any-time, and any-device, there is undeniably an era of Big Data” (Sivarajah et al. 2017, p. 263) Big Data is defined as “fast growing amount of data from various sources that increasingly poses a challenge to industrial organizations and also presents them with a complex range of valuable-use, storage and analysis issues” (Addo-Tenkorang & Helo 2016, p. 528). Two puzzle pieces have come together to shape Big Data: Large amounts of diverse data are available and the emergence of new analytical systems (Russom 2011). While at the beginning of this century the increase of digital data has been a problem for most information technology (IT) systems, today it is not a technological problem anymore but instead considered to be a competitive advantage (Tien 2013).

In the manufacturing industry, the concept of Big Data is not widely used yet and has, despite the fact that data-rich environments can be found in modern manufacturing companies, only recently been discovered (Addo-Tenkorang & Petri T Helo 2016; Rüßmann et al. 2015; O’Donovan et al. 2015; Yin & Kaynak 2015). On the shop floor high amounts of data in raw, unstructured format exists, which could, in the right way, be used to gain deeper insights into production operations (Lee et al. 2016). As production machinery can already generate large streams of data, Big Data usage is likely to increase, given the rise of other technologies such as IoT (Yin & Kaynak 2015). By combining the digital information from different sensors with the physical production processes, companies are able to create a high-performing manufacturing set-up (Sniderman et al. 2016). With falling prices for Big Data technology enabling such sensors, hardware and analytical software, this trend will grow even further (Manyika et al. 2011).

The business potential of Big Data is regarded as high (Chen et al. 2014). Today, Big Data are seen as an valuable asset like material assets or human capital (Chen et al. 2014). It is deemed to be the next frontier of innovation, competition and productivity (Manyika et al. 2011). Companies try to gain competitive advantage by exploiting the information richness of a variety of available data (Provost & Fawcett 2013). However, companies increasingly question how to derive value from this information explosion (Lavalle et al. 2011). However,
1 Introduction

1.2 Problematization

until Big Data are analysed to support the process of how business decisions are made, data are worthless (Tien 2013). Yet, it is difficult to utilize the flood of data and apply them to certain business situations in a useful manner (Russom 2011). Hence, the biggest challenges are not data themselves, but how to successfully implement and integrate Big Data analytics (BDA) in business operations (Lavalle et al. 2010). The analysis of Big Data is defined as “Data sets and analytical techniques in applications that are so large (from terabytes to exabytes) and complex (from sensor to social media data) that they require advanced and unique data storage, management, analysis and visualization technologies.” (Chen et al. 2012, p. 1166) With BDA companies are able to reveal insights that were hidden before to drive, manage and control the business (Russom 2011). The concept of BDA is closely connected and interwoven with Big Data and data-driven decision-making (Provost & Fawcett 2013). “Data-driven decision making (DDD) refers to the practice of basing decisions on the analysis of data rather than purely on intuition” (Provost & Fawcett 2013, p. 53). Studies have shown that companies, which embracing analytics are higher performing than those that do not (Lavalle et al. 2011; Brynjolfsson et al. 2011; Manyika et al. 2011).

In manufacturing, BDA creates new opportunities to improve productivity, efficiency and margins (Intel Corporation 2014). Acquiring production-related information is the basis for ‘data-driven operational excellence’ and highlights the importance of data as a valuable resource (Bechtold et al. 2014). The insights obtained from analytics can positively impact bottom-line business growth through improvements (Gjendem & Deep 2016). BDA will be the future competition lever to improve operations (Philip Chen & Zhang 2014). However, the current status in manufacturing companies is that large amounts of data are not used to improve production processes (Schröder 2016). This is due to the fact that large amounts of data are collected but are not used for analysis or cannot be accessed from storage (Koch et al. 2014). Although statistical data analysis was deployed by manufacturers before, the variety and richness of new Big Data demands new methods of analytics (Intel Corporation 2014). Thus, the exploitation of data value will be core in future manufacturing operations (Bechtold et al. 2014).

1.2 Problematization

Companies used to rely on their structured data which are stored in their internal databases to support the decision-making process (Dijcks 2013). Now, with new, unstructured data being collected from a range of sensors and sources outside of company boundaries, companies are seeking to combine multiple internal and external data sources with their already available internal data to drive decision-making (Dijcks 2013; Zhao et al. 2014). By combining new data sources with already available enterprise data, it can have a significant impact on the bottom line (Dijcks 2013). However, managers have to stay focused, as with the increased number of data sources and a high variety of data, it can become unclear which data are useful to drive business decisions (Lavalle et al. 2011). However, increased variety leads to more insights since the aim of data analysis is to derive at new insights utilizing the data from multiple sources (Davenport & Dyché 2013; Chong & Shi 2015). Thus, data need to be collected from a variety of sources to account for a certain degree of heterogeneity (Tien 2013). Yet, it becomes technically challenging to integrate data coming from many different data sources (Chong & Shi 2015). The main technical challenges are connectivity, where and which data to collect, how to integrate and analyse them (Lee et al. 2016). Nevertheless, Big Data goes beyond purely technological challenges (Abbasi et al. 2016). Challenges are not only related to data, but how to successfully integrate and manage BDA to improve business processes (Lavalle et al. 2010).

In contrast to companies operating on a digital platform such as Google, traditional companies such as manufacturing companies struggle in integrating BDA in their current infrastructure (Davenport & Dyché 2013). Thus, manufacturing companies face several
1 Introduction

1.3 Purpose

challenges and one of the future key capabilities will be integrated data collection and analysis (Koch et al. 2014). While some manufacturing companies have already made progress by launching first initial pilot projects, there are only a few who have implemented complete solutions (Blanchet & Rinn 2016). It is challenging to introduce advanced data analytics into daily operations (Lavalle et al. 2010), which is related to high uncertainty about the implementation requirements for Big Data as implementation results are mixed (McKinsey Digital 2016). Challenges currently hindering the implementation are lack of industry standards, data security and protection issues as well as the lack of qualified employees (Koch et al. 2014). Other challenges are the use of external data (Geissbauer et al. 2016), lack of courage (McKinsey Digital 2016), leadership (Davenport & Dyché 2013) coordination efforts to align analytics and business strategy (Lavalle et al. 2010), changing the decision-making culture and organizational layout (Brynjolfsson et al. 2011; Lavalle et al. 2011) as well as the limited capability of older IT systems (Intel Corporation 2014). The current status of manufacturing companies to use data as basis for value creation is poor as confirmed by a study from the consultancy Roland Berger: Roughly half of the companies in the study are not able to collect all data they potentially could, or the data is not stored and analysed (Rinn & Blanchet 2014). Case studies have shown that up to 99 per cent of all manufacturing data are lost in leakages (McKinsey Digital 2015). In a survey interviewing 1,500 companies about the current value of their data, on average only 14 per cent regarded data as valuable, 32 per cent as unneeded or insignificant and the remaining 54 per cent of data are from unknown content, termed ‘dark data’ (Ernst & Young 2016). Yet, to achieve a holistic view and total transparency of operations, data collection, integration and analysis from multiple sources are required (Bechtold et al. 2014). The overall sum of a large dataset is more valuable than analysing all different data on their own (Chen et al. 2014; Davenport & Dyché 2013). Therefore, one of the main challenges manufacturing companies face is the collection and integration of large amounts of data from multiple and widely spread internal and external data sources (Chen et al. 2014; Zhao et al. 2014).

Concluding, manufacturing companies recognize the clear need for strong data analysis capabilities, nonetheless only a minor share reported to have reached advanced levels (Geissbauer et al. 2016). Despite the fact that data have a significant value, only a few companies have the capabilities to exploit the value of Big Data so far (PricewaterhouseCoopers (PwC) 2015). This poses an imminent threat since based on the learning from the earlier industrial revolutions, companies who fail to change will not be able to compete in the long-term (Bechtold et al. 2014).

Based on the above described background information, the subsequent problem formulation can be formulated:

“With a large potential of various data types and sources waiting to be explored and exploited, manufacturing companies are not sure which capabilities are needed to successfully transform towards a data-driven organization utilizing internal as well as external data sources”

1.3 Purpose

This thesis’ purpose was to investigate which resources manufacturing companies have to acquire and which capabilities must be developed to overcome current challenges related to the implementation of Big Data analytics by utilizing internal and external data sources.

Due to the purpose of this investigation, this thesis focuses on the functional level. However, a system perspective was applied, as perspectives and aspects from the individual and industrial level were also considered.
1 Introduction
1.4 Research questions

1.4 Research questions

The following research questions have been formulated to operationalize the purpose of this research. The main research question (MRQ) can be answered with the help of the following two research questions (RQ).

RQ1: Which (new) internal and external data sources and types are considered by manufacturing companies for Big Data analysis?

RQ2: How can manufacturing companies successfully integrate data from internal and external data sources for Big Data Analysis?

a) Which challenges need to be overcome?

b) Which resources must be acquired and which capabilities developed?

MRQ: How can manufacturing companies explore and exploit internal and external data for Big Data analysis to transform towards a data-driven enterprise?

1.5 Research implications

The identified research gap is described and motivated as well as contributions are provided for academia and industry practitioners.

1.5.1 Rationale for research

Big Data is an emerging research area, which grew significantly over the last five years (Chong & Shi 2015). While earlier research has mainly focused on the technical site of Big Data (Chong & Shi 2015; Chen et al. 2016; Chen et al. 2014), recently management scholars have started to address the implications of Big Data on the firm and management level (Sivarajah et al. 2017). As Big Data covers several different areas such as information system, management and business research (Sivarajah et al. 2017), there is a need for research on Big Data and its implications on the wider organizational context and to adopt a holistic perspective on Big Data (Chen et al. 2016; Abbasi et al. 2016). The research field is still evolving, thus a full understanding of the topic and related aspects has not yet been established (Sivarajah et al. 2017). Thus, by contributing to Big Data research from a managerial perspective this thesis intends to minimize the research gap.

In the following, the research gap and rationale for this thesis' research are outlined along three aspects: The technical area of research, the theoretical frame of research and lastly the research methodology.

Technical area of research

The application of Big Data in the manufacturing sector is a relatively new research area, with growing interest since 2012 (O'Donovan et al. 2015). Since Big Data research is in its early stages, there are opportunities for multiple research avenues (Chen et al. 2016). Research regarding the actual implementation of Big Data projects is needed, especially in industries like manufacturing since marketing or customer-focused industries have already been studied (Bechtold et al. 2014; Chen et al. 2012; Provost & Fawcett 2013). Although the area of implementation research is developing now, it is still immature due to a chronological lag (O'Donovan et al. 2015). Especially the variety of Big Data is challenging, as integrating and analysing data originating from various data sources is technologically complex (Davenport & Dyché 2013). Consequently, integration is essential and it is thus worth studying how to integrate and manage different data sources for BDA (Chen et al. 2014).

Theoretical frame of research

Since Big Data incorporates the aspect of competitive advantage, the perspective of the ‘resource-based view of the firm’ (RBV) is an appropriate theoretical frame (Abbasi et al. 2016).
1 Introduction

1.5 Research implications

2016). Several studies, for example Gupta & George (2016), Akter et al. (2016), Côrte-Real et al. (2017) and Wamba et al. (2017), investigated Big Data capabilities using RBV. Although these capabilities have been empirically validated using quantitative surveys, Gupta & George (2016) stated that there is a clear absence of a holistic picture on Big Data capabilities and the list of resources which lead to BDA capabilities is yet not collectively exhaustive and is likely to expand over time. Moreover, as according to Phillips-Wren et al. (2015), “What capabilities (technical and non-technical) should an organization acquire to succeed in big data efforts?” is “arguably one of the most interesting questions in the field of big data research today” (Phillips-Wren et al. 2015, p. 465). Utilizing a process-focused research is supported by Maxwell et al. (2016) who stated that IT-related research should be best viewed along a process-perspective.

Methodology

In the recent systematic literature review by O’Donovan et al. (2015), much of Big Data research has been identified as being of philosophical nature. Scientific research mostly focused on theory and frameworks development and less on evaluation and validation (O’Donovan et al. 2015). However, it is crucial to develop theories which can be applied to problems in the field (O’Donovan et al. 2015). Moreover, the Fraunhofer Institute stressed the fact that for successful implementation experience in both academic research and practice is vital (van Ackeren & Schröder 2016). Likewise, Abbasi et al. (2016) emphasised the need for qualitative research in this area, exploring the interdependences between BDA, DDD and organizational transformation as well as cultural change. Alongside this, the majority of current research used quantitative surveys and hence little qualitative research exists (Sivarajah et al. 2017). In addition, a low level of case study research is prevalent due to low adoption of Big Data in organizations for now (Wamba et al. 2015). Therefore, these methodological avenues were selected for exploration in this thesis.

1.5.2 Originality and expected contribution

Due to the originality of the study, the research findings have implications for both researchers in the field of Big Data and for industry managers.

Originality

To the author’s knowledge this is one of the first studies that intends to map the resources and capabilities along a process perspective using a qualitative survey. Although other studies focusing on critical success factors and linking them to certain Big Data project phases have been conducted (Koronios et al. 2014), only general and non-industry specific recommendations were given. The phases were 1) business phase, 2) data phase, 3) analysis phase, 4) implementation phase, 5) measurement phase and 6) learning phase. They also applied the RBV but along the dimension of people, process and technology and identified the critical success factors for each stage and gave the main six CSF at the end. However, their research was built on secondary data generated from published case studies. Thus, future research is needed to verify these results with primary data (Koronios et al. 2014). What distinguishes the study by Koronios et al. (2014) from this thesis is the fact that they did not specifically emphasize data sources and data integration, instead they considered the implementation of Big Data from a holistic perspective. Likewise, the study by Vidgen (2014) focused on the challenges to overcome for each project phase using a multiple qualitative case study, however the setting did neither include the manufacturing sector nor was the study based on the RBV. Popović et al. (2016) conducted a multi-case study approach in the manufacturing sector and investigated relationships between Big Data capabilities and organizational readiness. Yet, they studied BDA usage levels from beginners to advanced but did not use a process perspective or focused on data source integration. Wamba et al. (2017) investigated Big Data capabilities applying a quantitative survey, but the study lacked aspects such as organizational culture, influence from top management and data security among others. Further, their empirical study was not aimed
1 Introduction

1.6 Disposition of thesis

at a specific industry. To conclude, although there are some studies investigating similar topics, this thesis distinguishes itself by applying a process-perspective to identify research and capabilities needed to overcome challenges on each process step. In addition, as mentioned before, increased focus is placed on the integration of different internal and external data sources since the integration poses major challenges.

Academic contribution
The academic contribution occurs on three aspects. First, a new perspective of Big Data resources and capabilities is provided by focusing on a process perspective. Second, the integration of data sources is of specific interest with new insights about the location of data source, the purpose of acquisition and finally their integration. Third, since Big Data implementation is investigated, a dynamic perspective is also provided indicating how manufacturing companies integrate Big Data sources over time.

Industry contribution
The industry contribution consists of strategic recommendations for guidance and support for managers in their efforts to implement Big Data in their organization.

Both academic and industry contribution are presented in detail in chapter 7.

1.5.3 Scope of research
This thesis’ research scope is delimited in four aspects:

Geographical delimitation
The geographical focus is Germany and Sweden (Central and Northern Europe) since interviews were only conducted in those countries due to reasons of accessibility of interviewees.

Time delimitation
The research was carried out over a 28-week research period between February and August 2017.

Industry delimitation
This study focuses on the manufacturing industry.

Big Data delimitation
This study is delimited to manufacturing companies, which have recently started first Big Data projects or undertook first initiatives in this manner.

1.6 Disposition of thesis
After this introductory chapter, this thesis proceeds as follows:

In chapter 2 an introduction to Big Data is given. It not only includes technological concepts of the technology enablers of Big Data in the manufacturing sector, but also a detailed overview of Big Data, its characteristics and data value chain.

In chapter 3 the theoretical frame is presented by outlining the RBV, the definition of resources and capabilities as well other related concepts such as dynamic capabilities. In the end, a conceptual research framework is presented, which serves as guidance and structure for the interview guidelines and results.
1 Introduction
1.6 Disposition of thesis

In chapter 4 the research methods are presented, described and motivated. The design, logic, process and outcome of the qualitative survey is described in all aspects and finally critically evaluated along four measures related to validity and reliability.

In chapter 5 the collected empirical material is presented by writing down the results and findings from the interviews. The results are analysed by utilizing the theoretical concepts presented in chapter 2 and 3.

In chapter 6 a discussion of this research findings follows to compare the findings with previous research. The findings for each research questions are discussed and finally a data archetype matrix is presented as output from this research.

In chapter 7 the research is summarized by reviewing the purpose and answering the research questions. Moreover, this research’s contributions to academia and practitioners are stated and final thoughts are shared on research limitations and future research.

Bibliography and appendix follow.
2 Introduction to Big Data in manufacturing

2.1 Enabling technologies for Big Data in manufacturing

This chapter provides background information related to Big Data, its characteristics and concepts as well as technological concepts related to the application of Big Data in the manufacturing sector. The aim of this chapter is to provide an overview of Big Data and to familiarize the reader with the most important concepts.

2.1 Enabling technologies for Big Data in manufacturing

In the following, a brief description of the technological concepts related to Big Data technologies in manufacturing is provided. The aim is to give an overview and familiarize with important concepts to understand this research better.

2.1.1 Industry 4.0

In 2011, the German government introduced the concept of ‘Industrie 4.0’, also known as the fourth industrial revolution as one out of ten future projects of the German High-Tech Strategy 2020 (MacDougall 2014). Increased digitization in the manufacturing sector led to the emerging of Industry 4.0 (Bitkom e.V. et al. 2016). Within this fourth industrial revolution the Internet has found its way into manufacturing, enabling more manufacturing intelligence and is the future vision of industrial production (Wang & Wang 2016). Industry 4.0 is a keystone to secure Germany’s technology leadership in the manufacturing sector (MacDougall 2014). While in the past manufacturing companies were focused on cost optimization and limited by the trade-off between profitability and asset turnover, Industry 4.0 solves these problems (Blanchet & Rinn 2016). Proposed benefits are more flexibility, reduced lead times, adapting customer demands and smaller batch sizes (Wang & Wang 2016). The total value effect of Industry 4.0 is estimated at EUR 420 billion and expectations are as high as 26 per cent productivity improvement (McKinsey Digital 2015; Blanchet & Rinn 2016).

However, Industry 4.0 is still perceived as a buzzword (Rinn & Blanchet 2014). Firstly, because Industry 4.0 is not a fixed defined term, making it hard to determine when Industry 4.0 is reached and secondly, because Industry 4.0 consist not of only one technology but a combination of innovations and variety of different technologies (Schröder 2016; Wang & Wang 2016). The novelty of Industry 4.0 is that a collection of new technologies such as Big Data, IoT and machine-to-machine communication (M2M), cloud technology, network communication and advanced analytics is enabling significant improvements (McKinsey Digital 2015; Bitkom e.V. et al. 2016). For an overview of these concepts and their relationships see figure 1. What impact those new technologies could have remains unclear and is still discussed by many companies (Rose et al. 2016).
2 Introduction to Big Data in manufacturing

2.1 Enabling technologies for Big Data in manufacturing

2.1.2 Internet of Things

Internet of Things (IoT) is a paradigm shift in building intelligent networks connecting machines one of the most disruptive technologies (Chen et al. 2014; Kees et al. 2015). IoT is defined as “connectivity of physical objects (things), equipped with sensors and actuators, to the Internet via data communication technology, enabling interaction with and/or among these objects” (Kees et al. 2015, p. 3). IoT allows communication via IP addresses where every object has a unique IP address (Wang & Wang 2016). Thus, IoT can be described as a network of smart objects, which can communicate with each other. IoT aims to digitalize operational activities by connecting and networking physical objects via the Internet (Rosemann 2013). The global trend for IoT is forecasted to be rapidly growing (Columbus 2016). For example, Gartner predicts 20.8 billion connected devices in 2020, up from 6.4 billion connected things in 2016 (van der Meulen 2015).

IoT in manufacturing companies

In industrial operations, the application of IoT devices and Big Data are going hand in hand to exploit their full potential (Addo-Tenkorang & Petri T Helo 2016). M2M communication technology as part of IoT are vital technologies for Industry 4.0 (Wang & Wang 2016). Thus, IoT is perhaps the most important enabling technology of Industry 4.0 (Sniderman et al. 2016). It is also described as the Industrial Internet of Things (IIoT) which connects all production assets, parts or products with sensors and embedded systems (Wang & Wang 2016). IoT sensors bridge the gap between the physical and virtual world and enable a never before achieved data transparency for improvement initiatives (Bechtold et al. 2014). Sensors attached to physical objects can monitor their status and send operating data to a centralized database (van Ackeren & Schröder 2016). These sensors are collecting data in order to analyse the information and extract new insights to control, regulate, alert, prevent or diagnose the production process (Bitkom e.V. et al. 2016). Collected data types are production data across all levels, machine operating data, process data and operator data (Intel Corporation 2014). They can consist of multiple data streams, coming from multiple locations and multiple floors (Fraunhofer IPT 2015). IoT sensors can be installed in the production are where currently data are lacking or cannot be captured in order to achieve full transparency of manufacturing operations (Koch et al. 2014).
2 Introduction to Big Data in manufacturing
2.1 Enabling technologies for Big Data in manufacturing

2.1.3 Cyber physical systems
Based on the technological foundation of IoT, connecting all relevant physical components within the manufacturing environment with IT systems via a network creates cyber physical systems (CPS) (McKinsey Digital 2015; MacDougall 2014). “A cyber physical system (CPS) is a complex system that integrates computation, communication, and physical processes” (Wang & Wang 2016, p. 1). The CPS is the core platform of Industry 4.0 (Schröder 2016). It consists of production equipment, tools, work pieces, plant and logistical components which are all linked to the Internet and communicate with each other via embedded software systems (Rinn & Blanchet 2014). Embedded systems have several mini-computers and sensors to capture all relevant information and data along the production process (Schröder 2016). These systems are further enhanced with modern communication interfaces to be networked via the Internet (Schröder 2016). CPS have two main functional requirements: connectivity technology to build a network between all physical objects and the virtual world as well as Big Data technology to manage, process and analyse data and subsequently use the insights to manage and adjust operations (Wang & Wang 2016; Lee et al. 2016).

CPS in manufacturing companies
Machine-to-Machine (M2M) communication allows automated information and data transfer between all parts of a CPS (Bechtold et al. 2014). Smart systems are created by connecting everything via the internet to form a network (MacDougall 2014). M2M communication technologies are mostly used in vertical integration, linking of all departments and production equipment, but will also be applied in horizontal integration, linking companies together (Bechtold et al. 2014). By connecting all relevant stakeholders in a network, constant data transfer enables the creation of dynamic value networks, which are operating in a self-organizing environment and can be optimized along different criteria such as costs, utilization or resource consumption (Bitkom e.V. et al. 2016). By combining the digital information from different sensors with the physical production processes, companies are able to create a high-performing manufacturing set-up, which provides advantages along time, quality, resources and costs in real-time (MacDougall 2014; Sniderman et al. 2016).

2.1.4 IT systems and Big Data platforms
The increasing amount of collected, stored and analysed data from the manufacturing environment requires new digital infrastructures (Bechtold et al. 2014). Current IT-systems such as Enterprise Resource Planning (ERP) or Manufacturing Execution System (MES) will face huge challenges, as they need to manage large amounts of data (MacDougall 2014). Thus, in order to benefit from the full data richness the CPS can collect, substantial IT investments in hard- and software are needed to cope with the large amounts of data and their analysis (Schröder 2016). New Big Data platforms are able to integrate and process structured data from existing databases (ERP, MES) but also unstructured data from new sensors. (Intel Corporation 2014). Big Data platforms are becoming of growing importance for traditional companies (Chong & Shi 2015).

Technologies that are related to Big data are cloud computing, master database management system(MDMS), Apache Hadoop (software processing of data) and MapReduce (new platform enables large scale data access) (Addo-Tenkorang & Petri T Helo 2016; Chen et al. 2012). As a result of new IT system introduction, the IT structure is getting more and more complex, due to upgrading older IT systems, new Big Data platforms and cloud technologies, leading to ‘coexistence’ between traditional IT systems and more advanced BDA technologies (Davenport & Dyché 2013; Abbasi et al. 2016).

Big Data platforms must have several characteristics to cope with the requirements of BDA (Davenport & Dyché 2013). At first, the fast and efficient processing of data requires the capability to integrate, manage and analyse the data using appropriate analytical tools
2 Introduction to Big Data in manufacturing

2.1 Enabling technologies for Big Data in manufacturing

(Davenport & Dyché 2013). The analysis of data also includes the visualization of insights (Chen et al. 2016). At second, the system infrastructure of the Big Data platform must have the capability to quickly capture, store and transfer large volumes of high variety data (Chen et al. 2016). Due to widely distributed Big Data sources, the platform should also have networking power to share data between different sub-systems (Chen et al. 2016). At last, the platform must be scalable and flexible in both data storage volume and computing power to be prepared to handle even more data in the future (Chen et al. 2016).

Cloud storage or cloud computing is another IT infrastructure part related to Big Data (Hu et al. 2014). Cloud consist of service offerings by providing infrastructure, platforms and software solutions (Assuncao et al. 2015). It is mostly used to store large amounts of data cost effectively and has the ability to be scaled up and down quickly in both data storage space and analytics computing power to meet a companies’ future demands of BDA (Hu et al. 2014; Addo-Tenkorang & Petri T. Helo 2016). Different types of cloud technologies exist, which mainly relate to location, accessibility and ownership (Assuncao et al. 2015). Clouds can be either purely private or public or a hybrid approach of both public and private.

Another type of IT technologies related to BDA are business intelligence systems, which aim to improve decision-making (Philip Chen & Zhang 2014). Business intelligence (BI) is “a business practice that measures and monitors the performance of business operations frequently.” (Russo 2011, p. 26). Today, almost all companies are relying on Business intelligence and analytics (BI&A) (Chen et al. 2012). BI&A is defined as “techniques, technologies, systems, practices, methodologies and applications that analyse critical business data to help an enterprise better understand its business and market and make timely business decisions.” (Chen et al. 2012, p. 1166). Although BI&A emerged in late 2000’s, Big Data and BDA have only recently emerged (Chen et al. 2012). In fact, BDA is significantly influencing BI (Russo 2011). While BI relies only on internal data, BDA relies on internal and external data (Kwon et al. 2014).

IT systems in manufacturing firms

Today, industrial manufacturing still needs manual support to a large extend (Fraunhofer IPT 2015). The ERP system is the main IT application for monitoring all business processes, while the MES governs the short-term planning and control process of the production (Schröder 2016). Traditional IT systems are not capable of to capture, store, process and analyse Big Data (Yin & Kaynak 2015). In order to make use of the collected data along the process, all IT-systems need to be networked in order to enable and ensure data exchange in real time across all departments (Schröder 2016). The vertical integration of IT systems to achieve a networked production system compromises all levels of IT applications, starting from sensors, controllers, production management systems up to the corporate planning (Bitkom e.V. et al. 2016). However, today’s status indicates that an overall integration of all IT systems is not yet reality (Rüßmann et al. 2015). Only a few companies so far have managed to connect all process steps and achieve constant data exchange between them (Fraunhofer IPT 2015). This is true for both, horizontal and vertical integration. Currently, sensors and machines in production are only partially networked and connected to each other and feed limited data into a manufacturing control system (Rüßmann et al. 2015). Thus, automated information exchange between planning systems, production systems and sensors is not yet achieved (Fraunhofer IPT 2015). In addition, the use of data storage in the cloud, using cloud computing for BDA and the migration of ERP systems towards clouds is only slowly picking up (Elragal 2014). Especially German companies are slow in adopting cloud technologies (Frank 2017). In a recent European study, only 15 per cent of German enterprises have yet adopted cloud technologies in contrast to 50 per cent of companies in cloud leader countries Finland and Sweden (Giannakouris & Smihily 2017). The reasons for hesitation regarding cloud technologies are mostly related to security and trust issues as well as lack of knowledge, loss of control and ownership (Khan & Malluhi 2010; Giannakouris & Smihily 2017).
2 Introduction to Big Data in manufacturing
2.1 Enabling technologies for Big Data in manufacturing

2.1.5 Manufacturing and Big Data
Fundamentally new within Industry 4.0 is the real-time availability of all relevant information (Bitkom e.V. et al. 2016). Within Industry 4.0, all parts of the manufacturing process are not only connected, but they communicate and analyse the collected data to use the retrieved information to drive the manufacturing process (Sniderman et al. 2016). By connecting all relevant stakeholders in a network, constant data transfer enables the creation of dynamic value networks, which are operating in a self-organizing environment and can be optimized along different criteria such as costs, utilization or resource consumption (Bitkom e.V. et al. 2016). In Industry 4.0, the collection of data is essential to control automated production processes and therefore data becomes an important manufacturing resource (van Ackeren & Schröder 2016). Hereby it is important to not only collect partially data, but use of more sensors and actuators to collect all data (Koch et al. 2014). Sensors can collect data internally (internal data integration) and using external interfaces to the environment (external data integration) (Rinn & Blanchet 2014). To move towards Industry 4.0, there is the clear recommendation to measure, collect and analyse all relevant process and sensor data (Koch et al. 2014). With Industry 4.0, companies are able to collect and evaluate data from a variety of different sources, such as production, ERP or customer management data as well as machine data related to energy consumption, material quality, time and cost (van Ackeren & Schröder 2016; Rüßmann et al. 2015).

Implementation of Big Data
The transition towards Industry 4.0 and Big Data poses a paradigm shift for manufacturing companies (Blanchet & Rinn 2016). In order to demonstrate the potential impact and gain practical experience with Big Data applications, companies need to launch lighthouse projects (Rose et al. 2016). As Big Data is key in Industry 4.0, reaping the value out of data is the basis for successful implementation (Geissbauer et al. 2016). To start with the implementation, the networking of all production equipment with software, IT systems and sensors which are collecting all process-relevant information as data is of high priority (Fraunhofer IPT 2015). There is fast forward in applying IoT sensors on manufacturing assets to collect data (Lee et al. 2016). Companies tend to retrofit their production plans by upgrading older equipment with new sensors (Rinn & Blanchet 2014). Only 40 to 50 per cent of the current manufacturing equipment needs to be replaced in Industry 4.0 compared to over 90 per cent in the third industrial revolution (McKinsey Digital 2015). As a result, there are likely not many greenfield approaches, where Industry 4.0 factories are built from scratch, rather retrofitting existing equipment and machines for lower cost (Rinn & Blanchet 2014). When applying Industry 4.0 on brownfield sites, the focus is on closing current data lacks to capture all relevant information across the production process (McKinsey Digital 2015). Due to the high variety of existing machinery on the shop floor from different manufacturers and from different ages, it is believed the retrofitting to automation software and sensors to be expensive (Intel Corporation 2014).

Status quo of implementation
Recent studies indicating that so far only few manufacturing companies have made substantial progresses in implementing first Big Data projects or initiatives to improve their operations. For example, Gartner reports that although many manufacturing companies are eager to use BDA and have already made first investments, only 15 percent of all surveyed companies are using them so far in production (van der Meulen 2016). The consultancy Boston Consulting Group studied the adopting status of 312 manufacturing companies in Germany in 2016: Only one-fifth of the observed company have implemented a first concept of Industry 4.0 yet (Lorenz et al. 2016). Findings from the McKinsey survey in January 2016 indicated that 91 per cent of the observed companies are expecting operational improvements in efficiency and effectiveness, but only 68 per cent feel prepared and only 56 per cent have made so far progress in implementing first Industry 4.0 concepts (McKinsey Digital 2016). However, many company have ambitious goals for the future: A study by PwC
with over 2,000 companies (21 per cent of them in the manufacturing sector) revealed that 83 per cent of companies expected themselves to use DDD and 75 per cent want to apply BDA in the next five years to improve manufacturing operations (Geissbauer et al. 2016).

2.2 Characteristics of Big Data

There is no unified definition of the term Big Data, despite a steadily growing research interest (Chong & Shi 2015). Many definitions exist around the fact that Big Data limits the capacity of traditional IT systems to conduct analysis (Cukier & Mayer-Schoenberger 2013). For example, Manyika et al. (2011) define Big Data as “datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyse” (Manyika et al. 2011, p.1). Similarly the definition by Chen et al. (2014) relates to “datasets that could not be perceived, acquired, managed, and processed by traditional IT and software/hardware tools within a tolerable time” (Chen et al. 2014, p. 173). Based on these examples, other definitions of Big Data incorporated the need of new technologies: “Big data technologies describe a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and/or analysis.” (Villars et al. 2011, p. 3). Big Data is created through the fact that basically everything can be digitally measured, captured and stored (Sivarajah et al. 2017).

Since this research addressed data sources and their associated challenges for organizations, the following definition by Addo-Tenkorang & Helo (2016) served as definition for Big Data: “fast growing amount of data from various sources that increasingly poses a challenge to industrial organizations and also presents them with a complex range of valuable-use, storage and analysis issues” (Addo-Tenkorang & Helo 2016, p. 528). This definition highlights several aspects, which are central to this research, including ‘fast growing amount of data’, ‘various sources’ and ‘challenges’ as well as ‘valuable-use, storage and analysis issues’. These aspects are described below.

2.2.1 The five Vs of Big Data

In today’s time, Big Data is, in the scientific and practical literature, often characterised by the five Vs of Big Data, which relate to Volume, Velocity, Variety, Value and Veracity (Wamba et al. 2015). In 2001, Doug Laney from the Meta Research Group introduced the first three Vs related to Big Data, namely Volume (depth/breadth of available data), Velocity (near real-time data collection and processing of created data) and Variety (many different data types and data sources) (Laney 2001). In addition, Russom (2011) defined for these three Vs the following characteristics, see figure 2.
The other two dimensions of Big Data, Value (discover unknown insights (Davenport & Dyché 2013) and Veracity (the uncertainty and unreliability of data (Gandomi & Haider 2015) were added later on (Yin & Kaynak 2015).

2.2.2 Big Data value chain
A possible visualization of Big Data is to view data as a supply chain: Data are generated from different sources, integrated and combined with other data, followed by being merged through the analytics to result in valuable insights from which actions can be taken to improve the overall efficiency of the company (Accenture Analytics 2014). The Big Data value chain describes the successive process from collecting data to exploiting the value of the data (Miller & Mork 2013).

Many researchers have formulated varying Big Data value chains. Tien (2013) proposed four steps for BDA, which are 1) data acquisition, 2) data access, 3) data analytics and 4) application. Yin & Kaynak (2015) added the step of data generation before data acquisition and introduced data storage prior to data analytics. Other researchers added even more steps, especially related to preparation of data before data are analysed. For example, Philip Chen & Zhang (2014) proposed the following process: 1) data recording, 2) data cleaning, integration and representation, 3) data analysis, 4) data visualization and interpretation and 5) decision making (Philip Chen & Zhang 2014). The 5V of Big Data can also be mapped onto a process perspective or value chain of Big Data (Addo-Tenkorang & Petri T Helo 2016). Hereby, acquisition equals Variety, storage represents Volume, analysis is Veracity, application relates to Velocity and value-adding describes Value of Big Data (Addo-Tenkorang & Petri T Helo 2016).

A comprehensive framework of the abovementioned process steps is provided by Miller & Mork (2013), as they formulated seven steps in the data value chain. Yet, they clustered these steps into the three overarching categories data discovery, data integration and data exploitation. Data discovery covers data collection, data preparation and data organization. Data integration consists only of one step (data integration), while data exploitation entails data analysis, data visualization and data-driven decision-making. For an overview of the data value chain, see figure 3.

Concluding, the data value chain “provides a framework with which to examine how to bring disparate data together in an organized fashion and create valuable information that can inform decision making at the enterprise level” (Miller & Mork 2013, p. 3). Therefore, this framework with its three steps serves as guidance for structuring this chapter.

2.2.3 Data discovery
In the digital era more and more data sources are becoming available and new technologies, such as embedded sensors and mobile devices, not only generate data but also open up
2 Introduction to Big Data in manufacturing

2.2 Characteristics of Big Data

more potential data value opportunities (Addo-Tenkorang & Petri T Helo 2016). However, capturing as many data as possible (volume) is not enough, as rather the variety makes the difference (Russom 2011). The overall sum of a large variety dataset is more valuable than analysing all different data at their own (Chen et al. 2014). New insights and discoveries can only be made, if new data sources are utilized better (Russom 2011).

Within the scholarly literature the following classifications of data sources and types can be found: Data can either be structured, semi-structured or unstructured data (Intel Corporation 2014; Chen et al. 2013). Structured data are mainly found in tables, spreadsheets or relational databases and follow certain standards, structures and labels, whereas semi-structured data do not conform with such standards (Gandomi & Haider 2015). Unstructured data encompasses all data that are not in a fixed format, such as texts, images, audio or video files (Gandomi & Haider 2015). Nowadays, both unstructured and semi-structured data can be generated by a single data source, leading to a high degree of heterogeneity of Big Data (Chen et al. 2013).

Data sources can be classified into internal and external data sources (Zhao et al. 2014). Internal data relates to all data which are stored in enterprise-specific systems and are generated by internal operations, such as sales, supply chain management, customer orders or transactions (Gandomi & Haider 2015). Internal data includes mostly historical and structured data (Chen et al. 2014). In contrast, external data are generated by sources that lie outside the organization's boundaries and consist of data coming from websites such as e-commerce platforms, sensors or mobile phones (Gandomi & Haider 2015). These kind of data “may not directly be related to the firm’s business operations, but can provide novel and more flexible perspectives” (Zhao et al. 2014, p. 172), whilst internal data variety is often limited (Zhao et al. 2014).

George et al. (2014) identified five sources of Big Data, which are: public and private data, data exhaust as well as community data and self-quantified data. Public data are “typically held by governments, governmental organizations, and local communities that can potentially be harnessed for wide-ranging business and management applications” (George et al. 2014, p. 322). These data describe transportation, health care or energy usage for example. Private data “are data held by private firms, non-profit organizations, and individuals that reflect private information that cannot readily be imputed from public sources” (George et al. 2014, p. 322). For instance, transaction data of companies or movement of goods through a supply chain. Data exhaust „refers to ambient data that are passively collected, non-core data with limited or zero value to the original data-collection partner. These data were collected for a different purpose, but can be recombined with other data sources to create new sources of value” (George et al. 2014, p. 322). Examples are mobile phone usage or Internet searches. Community data are „distillation of unstructured data – especially text – into dynamic networks that capture social trends” (George et al. 2014, p. 322). These data include consumer reviews or social network posts and activities. At last, self-quantification data are „data that are revealed by the individual through quantifying personal actions and behaviors” (George et al. 2014, p. 322). For example, wearable devices that generate data based on the individual’s movement.

A structured classification was based on a consolidation of different other authors’ work by Hartmann et al. (2016). They proposed eight different data sources which are divided into internal and external sources. Internal data can be further categorized into existing data, which are available already in the company’s internal IT systems but are not currently used and self-generated data, for example sensor data. Regarding external data, the authors distinguish between acquired or purchased data from data providers, business partner provided data (either to be provided by suppliers or customers) and freely available data which can be acquired for no cists, which can be either open data (pre-structured and
2 Introduction to Big Data in manufacturing

2.2 Characteristics of Big Data

downloadable), social media data (e.g. Twitter or Facebook) or web data in general. An overview of their classification scheme can be found in figure 4.

Figure 4: Data sources as identified by Hartmann et al. (2016), own visualization

In order to receive valuable and meaningful insights, companies have to utilize data from multiple sources (Accenture Analytics 2014). This also includes data sourcing from external sources to enhance internal available data (Accenture Analytics 2014; Abbasi et al. 2016). In fact, new data sources can be found mostly in the company’s environment, beyond their physical boundaries (Davenport & Dyché 2013). Yet, the question to be answered is: what other data sources can be leveraged? (Abbasi et al. 2016). To answer this question, the aspect of data value must be considered (Dijcks 2013). It is challenging to determine which data sources are valuable for data analysis since the value of data can be different across data sources (Dijcks 2013). In general, higher data value and new information can be found in new, unstructured data sources (Dijcks 2013).

Once it has been determined, which data sources are valuable, the next step is to enable access to these data sources (Miller & Mork 2013). Firstly, gaining access to these data sources is important, as otherwise missing data and therefore a lower value of the data might arise (Kwon et al. 2014). According to Klein & Verhulst (2017), there is confusion around data ownership and granting others access to data sources and issues and challenges remain. To gain access to data sources, technical as well as governance requirements have to be considered and related risks need to be managed by all involved parties (Klein & Verhulst 2017). Risks are legal aspects related to data privacy and data confidentiality as well as the necessary financial investments for data sharing infrastructure and possible staff training costs (Klein & Verhulst 2017). As a countermeasure to risks, it is proposed to put emphasis on mutual benefits and incentives like developing analytical capabilities, possible new revenue generation and boost the reputation as business partner (Klein & Verhulst 2017).

Data discovery in manufacturing companies

In manufacturing companies the volume and variety of data is also growing, which creates new opportunities to analyse data to improve productivity, efficiency and margins (Intel Corporation 2014). Traditional enterprise data from ERP and CRM systems (customer transactions, inventory management, sales management, financial data) and machine-generated data or sensor data but also social data (customers, social media) can be found in manufacturing companies (Dijcks 2013; Kambatla et al. 2014). Currently collected data types relate to production data across all levels, machine operating data, process data and operator data (Intel Corporation 2014). Acquiring production-related information is the basis for ‘data-driven operational excellence’ and underlines the importance of data as the valuable resource in Industry 4.0 (Bechtold et al. 2014). Hereby it is important to not only collect partially data, but use more sensors to collect all relevant data (Koch et al. 2014). Sensors can collect data internally (internal data integration) and using external interfaces (Rinn & Blanchet 2014). Also, sensor-based data (semi-structured) yield additional insights,
2 Introduction to Big Data in manufacturing

2.2 Characteristics of Big Data

Data sources can be stored in a structured and organized way. This integrating data. First, an appropriate data internal and external data items across different data sources integrating internal and external data they often do not come in a certain format nor are they co (Kwon et al. 2014). Since the integration of data poses challenges IT systems (Abbasi et al. 2016). Data quality has two dimensions, data consistency and data completeness respectively (Kwon et al. 2014). Data consistency refers to “keeping data uniform as they move across the network and are shared by various applications and systems within a company or between companies (e.g. supply chain)” (Kwon et al. 2014, p. 389). Data completeness describes the “degree to which all data necessary for current and future business activities (e.g. decision making) are available in the firm’s data repository.” (Kwon et al. 2014, p. 389).

When collecting data from many different sources and integrating them, it is difficult to maintain data quality (Kwon et al. 2014). Yet, it is important to have a thorough integration process and maintain data quality to be able to extract reliable information from the data (Kwon et al. 2014). Data quality has two dimensions, data consistency and data completeness respectively (Kwon et al. 2014). Data consistency refers to “keeping data uniform as they move across the network and are shared by various applications and systems within a company or between companies (e.g. supply chain)” (Kwon et al. 2014, p. 389). Data completeness describes the “degree to which all data necessary for current and future business activities (e.g. decision making) are available in the firm’s data repository.” (Kwon et al. 2014, p. 389).

Since the integration of data poses credibility and quality issues, it is difficult to perform data integration as it challenges IT systems (Abbasi et al. 2016). However, to develop and achieve BDA capabilities, firms “must integrate their internal and external data” (Gupta & George 2016, p. 1052). For successful use of internal and external data for analysis, Dijicks (2013) argued that companies need to invest and develop in their IT infrastructure to be able to handle the integration of external data and process them together with already existing internal data. In particular, data coming from external sources are difficult to integrate since they often do not come in a certain format nor are they complete or only consist of relevant data (Zhao et al. 2014). Zhao et al. (2014) pointed out two major challenges when integrating internal and external data (Zhao et al. 2014, p. 173): (1) connecting common data items across different data sources and (2) selecting the relevant data items from both internal and external data for analysis. Fisher et al. (2012) presented two critical steps when integrating data. First, an appropriate data architecture should be developed, where multiple data sources can be stored in a structured and organized way. This architecture should
follow an logical and efficient structure in order to reduce computation power requirements when processing these data and must be in place before first data sets are uploaded and integrated (Fisher et al. 2012). Secondly, once an architecture was chosen, the to be integrated data sets must be shaped to the structure of the architecture. Therefore, datasets “need to be organized, partitioned, and prepared before they are uploaded” (Fisher et al. 2012, p. 55).

Data integration in manufacturing companies
Only when a holistic data collection approach is adopted, compromising ERP, machine and operating data as well as MES and energy data, companies can obtain the desired information transparency to drive improvement initiatives (Fraunhofer IPT 2015). When combining new data sources with already available enterprise data, the bottom line can considerably be improved (Dijcks 2013). The holistic view and transparency of operations requires data integration from multiple sources (Bechtold et al. 2014). To achieve this, it is important that all relevant IT systems are networked and integrated (Fraunhofer IPT 2015). Therefore, in manufacturing one of the key capabilities for achieving operational efficiency and effectiveness will be to integrate data from all sources and eventually and analyse them (Koch et al. 2014; McKinsey Digital 2015). One main challenge is to connect all sensors and integrate their data streams into one single platform or application for analysis purposes (Bitkom e.V. et al. 2016). The challenge of data integration is even larger, when data flows occur on both horizontal (suppliers, customers) and vertical integration (across departments) (Schröder 2016). After data integration there is a need to analyse the data, otherwise it is without value (Chen et al. 2014).

2.2.5 Data exploitation
Only collecting and storing data without further processing or analysing does not bring value to the organization (Sivarajah et al. 2017). As Big Data is more complex due to the large variety of data, it is not possible to analyse them using traditional approaches (Yin & Kaynak 2015). There is a range of different new analytical methods, from BDA, text analytics, web analytics, network analytics and mobile analytics (Chen et al. 2012). The analytical part of Big Data is only a minor process step in extracting insights from data (Sivarajah et al. 2017). After data have been analysed, the results of the analysis have to be visualized. The task of data visualization is to present the data insights in an graphical way so individuals can easily understand the information and act accordingly (Davenport & Dyché 2013; Lavalle et al. 2010).

Until Big Data are analysed to support the decision-making process, data are worthless (Tien 2013). Value can be created when insights are used to make data-driven business decisions (Sivarajah et al. 2017; Tien 2013; Abbasi et al. 2016). Value can be retrieved not without human intervention, thus the management and business skills to integrate Big Data insights into daily business lag behind (Sivarajah et al. 2017). In order for data analytics to yield the expected value, the data strategy must be aligned with business strategy as well as organizational adaptions have to be made to effectively turn the gained insights into action (Lavalle et al. 2010). Applying a data perspective to the underlying business problems is key for effective BDA (Provost & Fawcett 2013). However, it is difficult to make use of the flood of data and apply them to certain business situations in a useful manner (Russom 2011). Crawford (2013) argued, the value of data is affected by humans, since they interpret these data and draw conclusions and relationships and thus introduce human biases. Therefore, not all data sets are purely objective.

When exploiting the information richness of a variety of available data, companies try to gain competitive advantage (Provost & Fawcett 2013). Companies who embrace real-time decisions from BDA will outperform those who do not (Villars et al. 2011). The wider the scope of using analytics for decision-making, the more likely companies perform well
2 Introduction to Big Data in manufacturing
2.2 Characteristics of Big Data

(Lavalle et al. 2011). Insights gained from analytics can positively impact bottom-line and foster business growth (Gjendem & Deep 2016). Vidgen (2014) formulated, inspired from the level hierarchy of IT systems from Venkatraman (1994) several stages of the progress of application and usage of BDA within a company. The model consists of six hierarchical stages where on each stage the scale and scope of applying the results of BDA to business decision increases. It starts in the fragmented stage, which limits the use of BDA to certain departments only, and continues over localized, functional, data-driven and evidence-based stages to the last stage, called essential stage, where all decisions are based on BDA. To conclude, BDA companies are able to reveal insights which were hidden before to drive, manage and control all business operations and decisions (Russom 2011).

Data exploitation in manufacturing companies
Some manufacturing equipment is generating large amounts of data flows, which traditional methods and IT systems are not capable of handling (Intel Corporation 2014). The rapid growth of data is outgrowing the technical advances made in IT systems limiting their ability to process these amounts of data calls for new methods of analytics (Intel Corporation 2014; Sivarajah et al. 2017). Data analysis covers a range of different methods such as statistics, data mining or machine-based learning to achieve continuous optimization (prescriptive analytics) (Bitkom e.V. et al. 2016). Still, only the collection of data does not mean to acquire new knowledge. It is therefore vital to transform the raw data into meaningful information (Fraunhofer IPT 2015). Data analysis is foremost performed to receive new insights and gain new knowledge (Bitkom e.V. et al. 2016). In order to exploit the information value from the collected data, manufacturing companies first have to understand where they are now and where they want to be (Sniderman et al. 2016). It is important to understand that only the data with the information most relevant to the desired vision will be of value for the company (Sniderman et al. 2016). The exploitation of data value will be core in future manufacturing operations (Bechtold et al. 2014). New insights can be used for the decision-making process to steer the production or drive general business actions (Bitkom e.V. et al. 2016). However, this implies the need for evaluation of all data along the production process (Rinn & Blanchet 2014).
3 Theoretical framework
3.1 Resource-based view of the firm

In this chapter, the theoretical frame of this thesis’ research is presented, described and motivated. Suitable theoretical frameworks and models are aligned with the course of investigation as chosen in the problematization. Therefore, this chapter considers theories regarding the resource based view of the firm, resources and capabilities as well as dynamic capabilities. At last, the theories are mapped to the topic of Big Data.

3.1 Resource-based view of the firm

This research is based on the resource-based view (RBV) of the firm, a well-known concept in management research and grounded theory (Barney 2001). In addition, the concept of dynamic capabilities (DC) is a useful tool in strategic management and provides a theoretical lens to investigate how an organizational deals with the uncertainties of innovation processes in changing environments (Teece et al. 1997; Eisenhardt & Martin 2000; Helfat & Peteraf 2003a; Lee & Kelley 2008; Ambriosini & Bowman 2009).

The resource-based view (RBV) of the firm was introduced by (Wernerfelt 1984), and developed by subsequent work of (Barney 1986; Barney 1991) as foundational work for grounded theory. Penrose (1959) made the initial work for the resource-based view of the firm by pointing out the new view of the firm from a resource perspective. Wernerfelt (1984) proposed “simple economic tools’ to “manage the firm’s resource position over time” (Wernerfelt 1984, p. 171). It serves as a complement to the industrial organization analysis introduced by Porter in 1980 (Porter 1980), however research started already in the 70’s with increased focus on the heterogeneity of resources and internal competencies of firms (Andrews 1971; Peteraf 1993).

3.1.1 Resources, capabilities and competencies

In the following, resources, capabilities and competencies will be defined and further explained. Nevertheless, the difference between resources, capabilities and competencies is not clear in scientific literature (Ambriosini et al. 2009). For example, (Barney 1991) defined the term ‘resource’ in a wider sense that it compromises not only resources but also activities, capabilities and competencies.

Resources

Amit & Schoemaker (1993) defined resources as “stocks of available factors that are owned or controlled by the firm” (Amit & Schoemaker 1993, p. 35). Other authors added more details, for example Helfat & Peteraf (2003) defined a resource as “an asset or input to production (tangible or intangible) that an organization owns, controls or has access to on a semi-permanent basis.” (Helfat & Peteraf 2003, p. 999). Teece et al. (1997) emphasized on the imitability aspect of resources when he defined resources as “firm-specific assets that are difficult if not impossible to imitate” (Teece et al. 1997, p. 516). A more general definition was given by Grant (1991), who simply stated: “Resources are inputs into the production process” (Grant 1991, p. 118). Since resources are also capabilities as has been outlined above, the following definition should hold for this thesis: “Resources are something that the organization can draw upon to accomplish its aims” (Helfat et al. 2007, p. 4).

Capabilities

Capabilities are defined as “a firm’s capacity to deploy resources, usually in combination, using organizational process to effect a desired end” (Amit & Schoemaker 1993, p. 35). Some capabilities need a complex set of different resources, while others just need one resource (Grant 1991). A capability is further defined as “the ability to perform a particular task or activity” (Helfat et al. 2007, p. 1) and are characterized as “complex, structured and
3 Theoretical framework

3.1 Resource-based view of the firm

"multidimensional" (Winter 2003, p. 992). The differences between resources and capabilities are outlined by Amit & Schoemaker (1993): "Capabilities [...] refer to a firm’s capacity to deploy Resources, usually in combination, using organizational processes, to effect a desired end. They are information-based, tangible or intangible processes that are firm-specific and are developed over time through complex interactions among the firm’s Resources." (Amit & Schoemaker 1993, p. 35). Thus, a capability distinguishes itself from a resource, since it is embedded in organizational processes and firm-specific (Makadok 2001). Therefore, a capability is defined as "a special type of resource – specifically, an organizationally embedded non-transferable firm-specific resource whose purpose is to improve the productivity of the other resources possessed by the firm" (Makadok 2001, p. 389). Further, capabilities are not just built up by a configuration of certain resources but they also need certain amount of coordination between the firm’s employees and other resources (Grant 1991). "Firm-specific assets are assembled in integrated clusters spanning individuals and groups so that they enable distinctive activities to be performed, these activities constitute organizational routines and processes" (Teece et al. 1997, p. 516).

Types of resources & capabilities

As outlined by Peteraf & Barney (2003), companies possess a varied mixture of resources and capabilities. Resources can be divided into tangible or intangible assets which are part of the firm (Wernerfelt 1984). Grant (1991) added the aspect of personnel resources. Thus, the resource base of the firm consists of tangible, intangible and human assets (resources) and additionally capabilities on a "preferential basis" (Helfat et al. 2007, p. 4) which the firm owns, controls or has access to. But owning is not mandatory to consider an asset as a resource (Helfat et al. 2007). Intangible resources can be patents, copyrights, brand names, trade secrets or reputation (Grant 1991). Tangible resources or assets refer to financial resources (capital, investments) and physical resources (technologies, machinery, plants) (Grant 1991; Wernerfelt 1984; Hofer & Schendel 1978). Human assets include knowledge and expertise, employee skills, relationships and leadership (Wernerfelt 1984; Barney 1991; Ross et al. 1996). Another widely used categorization of resources is to split them along four dimensions of financial resources, physical resources (machines, manufacturing facilities, buildings), human resources (experience, knowledge, skills) and organizational resources (relationships, culture, processes, routines) (Hofer & Schendel 1978; Barney 1995, p. 50).

Capabilities can be categorized in two dimensions according to Zollo & Winter (2002). Those are firstly operating routines, which are "geared towards the operational functioning of the firm" (Zollo & Winter 2002, p. 340) and DC, which modify the operating routines (Zollo & Winter 2002). Operational capabilities determine the status quo of the company, while DC describe the change of operational capabilities over time (Winter 2003).

Helfat & Peteraf (2003) introduced the capability lifecycle, which supports the explanation why there is heterogeneity of resources among firms. The cycle describes the evolution of both operational an DC consisting of the following stages (Helfat & Peteraf 2003b): founding stage, development stage and maturity stage which is then followed by six options: retirement (death of capability), retrenchment (gradual decline of capability), renewal (improvement of capability), replication (apply capability somewhere else), redeployment (take capability to somewhere else and adapt it slightly) and recombination (combine two or more capabilities) (Helfat & Peteraf 2003b). The six options can be simply segmented into either abandoning the resource or extend its life somewhere else (Helfat & Peteraf 2003b). The cycle is not seen as a process but rather evolving capabilities go through a number of different stages and can be altered substantially by doing so (Helfat & Peteraf 2003b).

3.1.2 Dynamic capabilities

Danneels (2002) argued that the RBV needs to have dynamic perspective to analyse how firms change their resources over time (Ambriosini et al. 2009). For example, capabilities
3 Theoretical framework

3.1 Resource-based view of the firm

and resources need to adapt to changing technology opportunities (Teece 2007). A firm has to develop its resource base to address ‘resource gaps’ in order to align the firm’s resource base to the current strategy (Grant 1991). Developing the resource stock does not only mean to build capabilities and resources internally but also means to acquire resources from external sources (Grant 1991). The concept of dynamic DC includes the capacity to identify the need or opportunity for change, formulate a response to such a need or opportunity, and implement a course of action (Helfat et al. 2007, p. 2). Thus, developing new capabilities helps firms to overcome challenges. Success and failure depend on whether an organization is capable of developing and utilizing DC (Teece 2007). Basically, if a company does not have DC, it will be successful only in the short run, but not for the long run due to change (Teece 2007). Therefore, a firm must develop DC to change and create new ways of value creation to stay ahead of the competition in the long run (Helfat et al. 2007).

Definition and descriptions

DC are defined as “the firm’s ability to integrate, build, and reconfigure internal and external competences” (Teece et al. 1997, p. 516) and as “the capacity of an organization to purposefully create, extend or modify its resource base” (Helfat et al. 2007, p. 1). The word dynamic represents the need for renew competences or resources to cope with change (Teece et al. 1997). While Teece et al. (1997) argue that DC are needed for firms which operate in “rapidly changing environments” (Teece et al. 1997, p. 516). However, other authors such as Zollo & Winter (2002) and Eisenhardt & Martin (2000) confirmed with their research that DC are also built when there is only a limited degree of change or even stable environments, thus there is no need for rapidly changing environments for DC to emerge. Thus, they defined DC as “a learned and stable pattern of collective activity through which the organization systematically generates and modifies its operating routines in pursuit of improved effectiveness” (Zollo & Winter 2002, p. 340). DC are deeply rooted within the firm’s resource base where they either deleting not needed resources anymore or reconfiguring the resource base (Sirmon & Hitt 2003). Further, since DC modify operational routines, it is based on learning of the organization (Zollo & Winter 2002). Therefore, DC can be also seen as organizational processes (Helfat et al. 2007). In fact, DC consist of four processes which are reconfiguration, leveraging, learning and integration (Teece et al. 1997; Bowman & Ambrosini 2003). Reconfiguration is the transformation and recombination of assets and resources. (Ambrosini et al. 2009) Leveraging is to extend or replicate a resource or capability across several domains (Ambrosini et al. 2009). Learning is to gain experience and to things faster and more efficient (Ambrosini et al. 2009). Finally, integration to coordinate resources to form new resource base (Ambrosini et al. 2009).

Types of dynamic capabilities

Within the scientific literature, several types and classifications of DC can be found. For example, Teece et al. (1997) proposed three categories of DC, namely processes, positions and paths.

Processes are either organizational or managerial and have three tasks: coordination/integration of resources in the static view of DC, learning in the dynamic perspective and finally reconfiguration as a transformational view (Teece et al. 1997). These tasks are a subset of three higher operating processes which are the “capacity (1) to sense and shape opportunities and threats, (2) to seize opportunities, and (3) to maintain competitiveness through enhancing, combining, protecting, and, when necessary, reconfiguring the business enterprise’s intangible and tangible assets.” (Teece 2007, p. 1319). These can be summarized as entrepreneurial management to identify problems and trends, decide on a course of action and subsequently align organizational resources and capabilities to the new strategy to capture value. The position of a firm is depending on the assets it has, which are: technological assets, complementary assets (to technology), financial assets, reputational assets (intangible) and structural assets, institutional assets, market structured and organizational boundaries (Teece et al. 1997). Lastly, DC are path
3 Theoretical framework

3.1 Resource-based view of the firm

dependent. Path dependency is described as “not only defines what choices are open to the firm today, but …also puts bounds around what its internal repertoire is likely to be in the future” (Teece et al. 1997, p. 515). Path dependencies limit a firm where it can go next depending on the current position which is influenced by past decisions made (Teece et al. 1997).

Wang & Ahmed (2007) also proposed three different DC, which they refer to adaptive capability, absorptive capability and innovative capability. Adaptive capability is the identification of emerging market opportunities and the ability to align the organizations resources to capture these opportunities. It includes all processes related to the management of organizational change in this matter and is a measure of strategic agility. Absorptive capability refers to the ability to recognize and understand the value of external information and to deploy the externally acquired knowledge to foster internal organizational learning (Cohen & Levinthal 1990). Absorptive capability can be further broken down into exchange capability and combinative capability (Zahra & George 2002). While exchange capability describes the process of evaluating and acquiring new knowledge externally, combinative capability terms the process of using the acquired external knowledge to enhance already available internal knowledge. These two capabilities can be further broken down into four sub-processes, which are the acquisition, assimilation, transformation and exploitation of new external knowledge by an organization (Cohen & Levinthal 1990; Zahra & George 2002). Innovative capability is the organizational ability to constantly be changing and innovative to develop new products, processes, services or routines (Wang & Ahmed 2007). It supports those organizational processes, which encourage innovation, change and development in the company.

Hierarchies of dynamic capabilities:

Besides different types of DC, the scientific literature also discusses hierarchies of DC. Several authors formulated possible hierarchies. For example, Collis (1994) developed a four level hierarchy consisting of 1) resource itself 2) modification of resource base 3) creation and extension of research base and 4) ‘Learning-to-learn’ capabilities. While Danneels (2002) only distinguished between first order and second order capabilities, Winter (2003) adapted this view slightly by adding zero-order capabilities (resource base, earn living in the present), first order capabilities (change in zero-order capabilities) and higher order capabilities (modifying firms DC). Based on the work by Collis (1994), Eisenhardt & Martin (2000), Danneels (2002) and Winter (2003), Ambrosini & Bowman (2009) proposed a new hierarchy consisting of incremental, renewing and regenerative DC. Incremental DC describe continuous, incremental changes to the resource base of the firm in stable environments (Ambrosini et al. 2009). Renewing DC relate to the refreshment and renewal of the resource base in changing environments (Ambrosini et al. 2009). At last, regenerative DC explain that certain DC need to be renewed, if the current DC cannot successfully alter the resource base to stay on top of competition (Ambrosini et al. 2009).

3.1.3 Competitive advantage

In RBV, firms are different to each other in their resource and capability base (Teece et al. 1997). Resources and capabilities are the basis for not only a firm’s strategy but also for profit earning of the firm (Grant 1991). Firms can create superior economic rents due to owning scarce firm-specific resources and gain hereby competitive advantage (Teece et al. 1997). Dierickx & Cool (1989) outlined the importance of internal resources to achieve competitive advantage. Resources are considered as being of competitive advantage to a firm, if they have the characteristics of value, rareness, imperfect imitability and substitutability (Barney 1991), which are now considered as the VRIN criteria (Ambrosini et al. 2009). Furthermore, resource heterogeneity influences competitive advantage (Conner 1991). Competitive advantage is a result of the process of changing the firm’s resources and capabilities (Teece et al. 1997). Hereby, resource picking and capability building are the two
3 Theoretical framework
3.2 Mapping the theory to Big Data

fundamental processes (Makadok 2001). Firstly, managers have to determine which resources are valuable to the firm in combination with the existing resources and secondly, managers have to be able to deploy these resources in an effective way (Makadok 2001). Capabilities cannot be acquired, they need to be built (Teece et al. 1997). Thus, managers are creating superior economic rents by building capabilities by combining and restructuring resources available to the firm (Makadok 2001). Managers define the business strategy by deciding where to build competences in the long term (Teece et al. 1997).

3.2 Mapping the theory to Big Data

The theoretical frame of RBV was used in previous research in relation to Big Data and is therefore applicable (e.g. Garmaki et al. 2016; Akter et al. 2016; Côrte-Real et al. 2017). In addition, Braganza et al. (2017) stated that RBV and DC provide each on their own only selective insights into the topic of Big Data, but together they provide a useful theoretical lens to investigate Big Data resources and capabilities. Furthermore, DC enable the investigation of how companies can implement Big Data in a repeatable and sustainable way by changing its resource and capability base (Braganza et al. 2017). Thus, the theoretical frame comprising RBV and DC is useful to address the purpose of this research by investigate which resources manufacturing companies have to acquire and which capabilities must be developed to overcome current challenges related to the implementation of BDA by utilizing internal and external data sources.

3.3 Big Data challenges, resources and capabilities

So far, numerous studies have applied the resource based view to the research topics of IT systems and Big Data. It all started in information systems research, when Liebermann & Montgomery (1988) introduced the concept of tangible and intangible IT resources. More recently, Kim et al. (2011) determined IT capability is a function of IT management capability as well as IT personnel capability and IT infrastructure capability. Recently, scholars have increased their attention to apply the RBV towards Big Data to identify the constructs which lead to the development of BDA capabilities of companies. Gupta & George (2016) differentiated between IT capability and BDA capability, which they define as “a firm’s ability to assemble, integrate, and deploy its big data-specific resources” (Gupta & George 2016, p. 1049). A firm’s BDA capabilities consist of a unique blend of financial, physical, human and organizational resources (Gupta & George 2016). In relation to this, the authors Gupta & George (2016) argued as follows: While mostly technical challenges can be overcome with technology progress and the emergence of new technologies, companies must also think about developing their human and intangible resources to overcome all challenges.

Hence, scientific literature, which applied the RBV to develop BDA capabilities, was screened to identify challenges, resources and capabilities. The following studies were used to cluster challenges, resources and capabilities related to Big Data: Byers 2015; Fosso Wamba et al. 2015; Sivarajah et al. 2017; Koronios et al. 2014; Wamba et al. 2017; Vidgen 2014; Garmaki et al. 2016; Akter et al. 2016; Gupta & George 2016; Bakshi et al. 2014; Côrte-Real et al. 2017; Braganza et al. 2017; Brynjolfsson & Hitt 2000; Dixon et al. 2010; Côrte-Real et al. 2014; Côrte-real et al. 2014; Erevelles et al. 2016; Gunasekaran et al. 2017; Kwon et al. 2014; Lavalle et al. 2011; Zicari 2014 and Russom 2013.

Challenges

When transforming into a Big Data driven organization, challenges arise on technical, architecture, skill, leadership and organizational level (Davenport & Dyché 2013). After reviewing the literature, challenges can be structured along these four levels, namely technological, managerial, individual and organizational challenges.
3 Theoretical framework

3.4 Conceptual research framework

Table 1: Big Data challenges

<table>
<thead>
<tr>
<th>Level</th>
<th>Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technological</td>
<td>Limited IT infrastructure, processing of data, data policies, access to data, integrating data, data quality, data security</td>
</tr>
<tr>
<td>Managerial</td>
<td>Data-business strategy alignment, leadership, managerial support, data privacy and governance, investment priorities</td>
</tr>
<tr>
<td>Individual</td>
<td>Lack of skills and competencies, education and training</td>
</tr>
<tr>
<td>Organizational</td>
<td>Organizational processes, organizational change, talent management, cultural change</td>
</tr>
</tbody>
</table>

Proposed Big Data resources and capabilities

In a similar fashion to challenges, the scientific literature proposed also several resources and capabilities, which lead to the development of BDA capabilities. The identified resources and capabilities are structured along the three dimensions of tangible resources (physical and financial), human resources (managerial and individual) as well intangible resources (organizational) to match the dimensions of challenges.

Table 2: Big Data resources and capabilities

<table>
<thead>
<tr>
<th>Level</th>
<th>Resources and capabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tangible</td>
<td>Infrastructure: New IT infrastructure in both soft- and hardware, analytical platforms which are compatible, flexible and modular, data storage</td>
</tr>
<tr>
<td></td>
<td>Data: Investments and capital, clear business value identified</td>
</tr>
<tr>
<td></td>
<td>Financial: Data strategy definition, top management support, managerial leadership skills such as coordination, control and planning</td>
</tr>
<tr>
<td>Human</td>
<td>Managerial: new competencies and skills related to data science, acknowledging data value, process knowledge, business knowledge and technical knowledge, relational knowledge</td>
</tr>
<tr>
<td></td>
<td>Individual: Data-driven culture, decision-making culture, multi-disciplinary teams, organizational learning and knowledge management, creative thinking</td>
</tr>
<tr>
<td>Intangible</td>
<td>Organizational</td>
</tr>
</tbody>
</table>

3.4 Conceptual research framework

Based on the information provided above, frameworks and theoretical concepts introduced and explained in chapter 2 and chapter 3, a conceptual research framework was created to guide this thesis’ research. See figure 5 for a visualization of the framework. The framework consists of two main perspectives, one perspective related to the Big Data perspective and the other related to the theory of the RBV of the firm. The Big Data perspective is structured along the analytical process of Big Data, as described by Miller & Mork (2013). For each process step, the challenges are investigated and possible resources and capabilities to overcome them are investigated on the RBV perspective. The research along a process is supported by Côrte-real et al. (2014) who stated that IT-related research should be best viewed along a process-perspective. A process perspective is important, since the sequence and coordination of the implementation of Big Data is important to consider (Maxwell et al. 2016). For example, when looking at the data value chain by Miller & Mork (2013), even if firms can determine the right data sources and successfully extract new insights from data, they can still fail in applying these new insights. Therefore, an end-to-end process...
3 Theoretical framework
3.4 Conceptual research framework

A comprehensive perspective is needed to cover all aspects of successful implementation of a Big Data project (Maxwell et al. 2016).

Figure 5: Conceptual research framework
In this chapter the process on how the data both for the scoping review and qualitative survey was acquired is outlined. After the research design is presented, sampling strategy, data collection and analysis are described, motivated and critically discussed. At last, the research methodology is critically evaluated with regards to validity and reliability measures.

4.1 Primary and secondary sources

Which sources or data are regarded as primary or secondary sources depends entirely on the investigation (Blomkvist & Hallin 2015). Since to the author’s knowledge no research has been conducted in this explicit research area, all literature compromising industry reports and scientific articles identified through a scoping review can be regarded as secondary sources, while qualitative, semi-structured interviews are considered as primary sources. The methodology of the scoping review will be outlined first as it lays the foundation for the qualitative interviews.

4.2 Scoping review

The purpose of the scoping review was twofold (Collis & Hussey 2014): Firstly, the review aimed at identifying previous studies in the research area to review the current body of knowledge and develop a broad view of the underlying phenomenon. This was beneficial for critically discussing the literature and detecting the research gap. Secondly, the literature review also facilitated the construction of the research design and build the data gathering process. To adequately perform a scoping review the five stage process for scoping reviews as proposed by Arksey & O’Malley (2005) was followed:

1. Identifying the research questions
2. Identifying relevant studies
3. Study selection
4. Charting the data
5. Collating, summarizing and reporting the results

4.3 Qualitative survey

Next, the method for collecting primary data from a qualitative survey are presented, described and motivated. It follows the structure of the four building blocks of design, logic, process and outcome of the research as proposed by Blomkvist & Hallin (2015).

4.3.1 Research design & purpose

A qualitative survey was chosen as study design, which intends to establish ‘diversity’ among the research topic (Jansen 2010). The systematic inclusion of diverse study participants is advantageous in this context since the research area represents an under-researched and dynamic field (Jansen 2010; O’Donovan et al. 2015). Due to the absence of explicit research and frameworks for this specific context, the purpose of this research is of exploratory nature (Blomkvist & Hallin 2015; Collis & Hussey 2014). Qualitative research aims to seek answers to questions of How? and Why? and is therefore considered appropriate to explore a phenomenon in detail and depth (Yin 2009; Collis & Hussey 2014; Anderson 2010). Since, in general, qualitative research methods are appropriate for the exploration of insufficiently studied topics, semi-structured interviews were performed to capture a range of resources and capabilities required to overcome Big Data challenges (Yin 2009; Collis & Hussey 2014). Some of the interviews have been conducted with employees.
4 Methods
4.3 Qualitative survey

at manufacturing companies, hence serving as case studies for this research. Case study interviews provide an excellent opportunity to investigate a phenomenon in its context (Easterby-Smith et al. 2012). However, due to limited access to case companies, the range of case studies might be too broad and their context is not controlled. Therefore, this research design is set to be a qualitative survey in the manufacturing sector with varying context to account for a broad coverage of the research topic and capture its full range.

4.3.2 Logic
This study will follow an abductive approach for two reasons: First, as aforementioned, there is lack of research, making it impossible to draw upon established theoretical frameworks. Second, due to the exploratory nature of the research, a solely inductive approach was regarded as inappropriate. Yet, there are critical considerations to be made. As a literature review was conducted prior to the empirical study, the likelihood of introducing bias exists, which influences a purely inductive approach. A high chance of influence from the reviewed literature and theory is present, which describes more the concept of a deductive research approach (Blomkvist & Hallin 2015). Therefore, this research can be categorized in between deductive and inductive research, thus an abductive approach (Blomkvist & Hallin 2015). As for the logic of this kind research, it also involves a continuous revision of the problematization, purpose and research questions if new insights are accumulated leading to another research direction (Blomkvist & Hallin 2015).

4.3.3 Process
The process describes the steps taken to acquire and analyse primary data from interviews.

Interview methodology
Semi-structured qualitative interviews were used to collect primary data sources. Semi-structured interviews consist of a predetermined structure or interview guide, but leave room to go more into detail, if new aspects emerge during the interview (Collis & Hussey 2014). This allows to investigate new dimensions of the phenomenon and creates flexibility if necessary (Blomkvist & Hallin 2015). However, this flexibility also hampers the comparison of interviews, as each interview differs in structure and content (Collis & Hussey 2014). Nonetheless, qualitative interviews are advantageous, as it simplifies the investigation of a complex problem and ask clarification questions (Blomkvist & Hallin 2015).

Question design
Data collection was supported by an interview guide (see Appendix A) that consisted of both open-ended and closed questions (Collis & Hussey 2014). While closed question were mainly used for employee-specific questions related to the role and previous experience as well for general questions related to the company, open-ended questions were helpful to learn about the interviewee’s opinions and experiences (Yin 2009). Hence, open-ended questions were asked, when it was required to understand a certain situation better, to get to know the experience of the employee in this particular situation and reflections of the actions made (Blomkvist & Hallin 2015). The interview guide contained all predetermined questions in a structured manner, covering the following areas and ranging from broad to more specific questions:

1. General information about the employee and company
2. Questions about IoT/ Big Data
3. Questions about data – sources, types, collection
4. Questions about data analysis – methods, process, insights
5. Questions about challenges along the process
6. Questions about resources and capabilities – main part
7. Wrap-up and closing
4 Methods
4.3 Qualitative survey

Sampling strategy
The targeted population of interviewees can be divided into three areas: The first population consists of employees from companies that provide software or services for BDA. This group is referred to as ‘Technology suppliers’. The Technology suppliers were expected to answer the research questions about data, data sources and the process of BDA. Further, if the companies provide some advisory services for other companies, they could also reveal some insights regarding resources and capabilities. The second population includes employees from manufacturing firms, which were involved in recent Big Data project. Those participants could share insights into the implementation process of the Big Data project and answered questions about resources and capabilities needed along the process. The employees’ companies are referred to as the ‘Case companies’, since the unit of analysis in this study is the implementation of a Big Data project on the firm-level (Yin 2009). These interviews provided the firm-perspective of Big Data implementation and gave insight to challenges, resources and capabilities in all organizational perspectives. The last study population is referred to as “Industry experts”, which all work in a major management consultancy firm and advise clients in the field of the research topic. This sub-population was interviewed last to validate already analysed findings from previous interviews and to close potential logical gaps. The selection of interviewees was guided by sampling principles as proposed by Eisenhardt (1989). The sampling strategy aimed to identify the key experts, which are very familiar with the research topic.

Regarding the ‘Technology suppliers’, an Internet search yielded 135 possible companies to interview. The search engine Google was used to identify these companies, using the following keywords: ‘Big Data analytics firm’, ‘Big Data service provider’, ‘Big Data start-up’, ‘Big Data implementation service’, ‘Big Data consulting’, ‘Data analytic company’, ‘Industrial Big Data company’, ‘Industry 4.0 supplier’, ‘Industry 4.0 company’. Inclusion criteria were either offering products or provide solutions within BDA and/or IoT sensors. Another inclusion criteria was to operate in the manufacturing sector. After those firms were identified, judgment sampling was performed to select the companies deemed to provide the best insights (Perla & Provost 2012). Hence, priority levels were allocated to each firm, ranging from the levels 1 to 3:

- **30 level-1-firms**: Headquarter and core market in Germany and provide products or solutions explicitly for the manufacturing sector
- **45 level-2-firms**: European focus and should serve the manufacturing sector among a wider industry portfolio
- **60 level-3-firms**: All other firms outside of Europe and with services or products for a wide range of industries

At first, firms with priority level 2 and some companies from level 3 were contacted via E-mail. While in most cases the corporate E-mail address was used, some companies provided the E-mail address from a contact person, which was then used to send out the E-mail. The companies contacted via E-mail received a compassing E-mail in either English or German (or both) with information regarding the overall topic, research focus, interview information as well as data handling and confidentiality. As additional information, an ‘Interview Teaser’ in PDF format was attached to the E-mail, providing additional information. The Interview Teaser was built in Microsoft PowerPoint® and consisted of 7 slides in total. After the first 50 E-mails were sent out, the E-mail text was reviewed and slightly changed, adding two more sentences to customize the E-mail to each specific company by indicating why this company can support this research by sharing its knowledge and experience. The refined version of the text was then sent to companies of level 1 priority, in hope to yield more positive answers. When a company did not respond to the E-mail within seven working days, a reminder E-mail was sent in hope to increase the likelihood of response (Meho 2006).
4 Methods

4.3 Qualitative survey

Regarding the identification of possible ‘Case companies’, newspaper reports and corporate press articles were searched for recent news about launched Big Data projects. Especially the Google News search was beneficial in identifying suitable companies, when searching with the following terms such as ‘recent ‘Big Data project’, ‘Big Data pilot project’, ‘Big Data lighthouse project’, ‘Big Data project launched’, ‘Industry 4.0 project started’. Inclusion criteria were: Operating in the manufacturing industry as primary industry, located in Germany or Sweden and within the last two years have started Big Data projects. The geographical scope of Germany and Sweden was used as part of a convenience sample strategy, as it was expected that given the research background and the thesis context in these two countries were the highest chances of finding companies that want to take part in this study (Anderson 2010). Although this kind of sampling strategy might introduce bias, it is regarded as appropriate for exploratory studies (Anderson 2010). The criterion that manufacturing companies must recently have started Big Data projects is based on a purposive sampling strategy (Anderson 2010). Companies fulfilling this criterion might be the most informative when it comes to Big Data implementation, as the actions taken are still vivid and thus can be recalled (Anderson 2010). In total, 64 manufacturing companies were identified from which 44 were contacted by a customized E-mail to ask for permission to interview.

Regarding the ‘Industry experts’, a convenience sampling strategy was applied to identify and contact knowledgeable industry experts from a leading management consultancy in Germany. The interviewees were selected based on their knowledge and experience in the fields of Big Data in manufacturing or Industry 4.0 and their timely availability.

Setting and data collection

Due to the geographical scope of the sample population as well as time constraints, all interviews were conducted via web meeting or telephone. While this set-up for data collection proofed to be fast, inexpensive and efficient to schedule interview dates, it was not possible to make observations and pick up non-verbal communication aspects such as body language, gestures and mimics (Collis & Hussey 2014). However, research has shown that there is no significant difference between telephone and face-to-face interviews (Sturges & Hanrahan 2004). All interviews were conducted within the limits of research ethics in terms of the interviews being voluntary, anonymous and interviewees could opt out at any given point or deny to answer a question (Blomkvist 2016). Prior to the interview, the study’s context and purpose were outlined and it was pointed out that the obtained information is limited is only used for this study. In the beginning, permission to record the interview was asked and in case of denial hand-written notes were made instead. Although it is suggested that qualitative research is content with a smaller sample size due to the workload associated with data analysis (Anderson 2010), the goal was to conduct as many interviews as possible until data saturation was reached to contribute to the research’s validity and reliability. Data saturation was defined as the point, when three consecutive interviews did not yield any new insights (Guest et al. 2006). Research suggests that data saturation in qualitative research is in general achieved after twelve interviews when the population is homogenous (Guest et al. 2006). However, given the heterogeneity of the population in this research, a larger number of interviews are needed and therefore this data saturation measure has been put in place, when no new data and no new coding themes could be retrieved from the interview transcripts anymore (Guest et al. 2006; Fusch & Ness 2015).

Qualitative data analysis

The audio recordings or taking hand-written notes enabled to transcribe each interview for the data analysis. This research followed a hybrid coding approach, using both inductive and deductive thematic coding procedures (Fereday & Muir-Cochran 2006). Thematic analysis results in emerging themes from the data, which are regarded as important for the research topic (Vaismoradi et al. 2013). By reading through the interview transcripts repeatedly, pattern recognition lead to themes that become the coding categories (Fereday & Muir-
4 Methods
4.3 Qualitative survey

Cochran 2006). By applying a hybrid coding approach, codes were either data-driven (inductive) or theory-driven (deductive) developed (Boyatzis 1998; Crabtree & Miller 1999). This approach allowed both deductive codes from theory and emerging codes from the data. The deductive approach is also known as template approach, where a research framework dictates the codes before data are analysed (Crabtree & Miller 1999). In this research, the codes were derived from the conceptual research framework, which was based on grounded theory and the research questions.

The process of data analysis was as follows: At first, all interview transcripts were read through multiple times to become familiar with the content. Secondly, each information or sentences, which related to the research framework, were highlighted in the text and subsequently summarized into shorter phrases. Thirdly, all phases were interpreted to identify their possible meaning and then given codes related to the meaning. At last, all codes were re-organized and structured. Moreover, it ensured that the categories and codes were linking. Throughout the whole process, the interview transcripts were read repeatedly to guarantee that the codes and categories were still close to the interviews and did not lose their relation to the information provided by the interviewees. Table 3 provides an overview of the final themes and codes indicating also whether it was deductive or inductive coding. The results were structured primarily along the research questions and secondarily the research framework as presented in chapter 3.4.

Table 3: identified themes and codes of data analysis

<table>
<thead>
<tr>
<th>Research question</th>
<th>Theme</th>
<th>Codes</th>
<th>Type of coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>RQ1</td>
<td>New data sources</td>
<td>- Area of data collection and data types</td>
<td>Deductive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Purpose of data source acquisition</td>
<td>Inductive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Data source integration over time</td>
<td>Inductive</td>
</tr>
<tr>
<td>RQ2</td>
<td>Big Data value chain</td>
<td>- Data discovery</td>
<td>Deductive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Data integration</td>
<td>Deductive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Data exploitation</td>
<td>Deductive</td>
</tr>
<tr>
<td>RQ2 a</td>
<td>Challenges</td>
<td>- Technical challenges</td>
<td>Deductive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Managerial challenges</td>
<td>Deductive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Individual challenges</td>
<td>Deductive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Organizational challenges</td>
<td>Deductive</td>
</tr>
<tr>
<td>RQ2 b</td>
<td>Resources &amp; capabilities</td>
<td>- Technical capabilities</td>
<td>Deductive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Managerial capabilities</td>
<td>Deductive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Individual capabilities</td>
<td>Deductive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Organizational capabilities</td>
<td>Deductive</td>
</tr>
</tbody>
</table>

4.3.4 Outcome

Overall, the research outcomes were two-fold. First of all, this research can be considered as applied research, since problems are discussed in a company-related environment (Collis & Hussey 2014). Thus, practical implications and managerial recommendations are outlined in the last chapter. However, contributions to academia are also made by discussing the research findings in comparison to previous studies and highlighting differences as well as new perspectives.
4 Methods

4.4 Overview of interviews

In total, 27 interviews have been conducted with a total of 1,395 minutes of interviewing time. In the following, descriptive statistics about the empirical material regarding the response rates and dropout analysis are presented as well as a detailed overview of all interviewees for each sample group.

4.4.1 Response rates and dropout analysis

Table 4 provides an overview of the response rates and number of interviews for each sample group.

<table>
<thead>
<tr>
<th>Sample Group</th>
<th>Abs.</th>
<th>%</th>
<th>Abs.</th>
<th>%</th>
<th>Abs.</th>
<th>%</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology suppliers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identified</td>
<td>135</td>
<td>n.a.</td>
<td>64</td>
<td>n.a.</td>
<td>5</td>
<td>n.a.</td>
<td>204</td>
</tr>
<tr>
<td>Contacted</td>
<td>127</td>
<td>94%</td>
<td>44</td>
<td>69%</td>
<td>5</td>
<td>100%</td>
<td>176</td>
</tr>
<tr>
<td>Responses</td>
<td>35</td>
<td>28%</td>
<td>10</td>
<td>23%</td>
<td>5</td>
<td>100%</td>
<td>50</td>
</tr>
<tr>
<td>Interviews</td>
<td>17</td>
<td>13%</td>
<td>5</td>
<td>11%</td>
<td>5</td>
<td>100%</td>
<td>27</td>
</tr>
<tr>
<td>Manufacturing companies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identified</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contacted</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Responses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interviews</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Industry experts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identified</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contacted</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Responses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interviews</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For further information about the technology suppliers, table 5 provides a more detailed view of the response rates for each priority level:

<table>
<thead>
<tr>
<th>Priority Level</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identified</td>
<td>30</td>
<td>45</td>
<td>52</td>
<td>135</td>
</tr>
<tr>
<td>Contacted</td>
<td>100%</td>
<td>100%</td>
<td>87%</td>
<td>94%</td>
</tr>
<tr>
<td>Responses</td>
<td>23%</td>
<td>27%</td>
<td>31%</td>
<td>28%</td>
</tr>
<tr>
<td>Interviews</td>
<td>17%</td>
<td>11%</td>
<td>13%</td>
<td>13%</td>
</tr>
</tbody>
</table>

The overall response rate of 28 per cent resulting in 27 interviews (15 per cent positive response rate) can be explained by the following reasons: Firstly, the E-mail was simply not read or was deleted before reading due to information overload in corporate E-mail accounts (Meho 2006). Secondly, the E-mail was read but due to the specific research topic and potential confidentiality issues it was decided not to engage in this research without further notice. Thirdly, by E-mails being sent to corporate addresses it might happen that the E-mail was not forwarded to relevant people regarding the topic of Big Data. Fourthly, the E-mail subject line included the word ‘interview’, which might be seen as a request for support or help and thus might have been perceived as spam (Porter & Whitcomb 2005).

There might be other reasons for the low response rate, but the final number of 27 interviews surpassed the suggested number of interviews until data saturation is reached (Guest et al. 2006). Possible strategies to enhance the response rate, for example to establish contact via telephone first before sending an E-mail with further details, were not executed due to time constraints (Kiezebrink et al. 2009). However, as aforementioned, a reminder E-mail was sent to non-respondents after one week, which led to a higher response rate (Meho 2006). In addition, by offering potential interviewees the option of either a telephone interview or interview via web meeting, it was assumed that participation rates increase due to a high flexibility by providing suitable situations for interviewees (Janghorban et al. 2014). Fifthly, the targeted sample of industry experts and practitioners was more difficult to motivate compared to academics, who were more open towards supporting empirical research (Sappleton & Lourenco 2016).
Nevertheless, as can be seen from the tables above that not all E-mail responses led to the establishment of an interview. Out of 50 responses, 23 responses did not yield an interview. While some respondents simply stated that they do not have the resources of both suitable employees to interview and time to participate in this research, others did not provide a specific reason. The exact circumstances or possible influences of why respondents turned down invitations are not clear and probably depend on the individual (Sappleton & Lourenco 2016). From the few studies that exist about this topic, the absence of financial incentives, complexity, effort, confidentiality and possible threats to the anonymity of the interviewee have been identified as possible factors to deny such interview requests (O’Neil et al. 2003). Some of these listed factors may be applicable here, for example most likely the lack of incentives and due to the novelty of the topic leading to confidentiality concerns.

### 4.4.2 Technology suppliers

From the technology supplier perspective, a total of 17 qualitative, semi-structured interviews were conducted. Details can be found in table 6 below.

<table>
<thead>
<tr>
<th>ID</th>
<th>Company</th>
<th>Role</th>
<th>Language</th>
<th>Date</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>TS1</td>
<td>IoT provider</td>
<td>Sales Engineer for Big Data solutions</td>
<td>English</td>
<td>2017-03-16</td>
<td>75 minutes</td>
</tr>
<tr>
<td>TS2</td>
<td>Big Data analytics provider</td>
<td>Data Scientist for production optimization</td>
<td>English</td>
<td>2017-03-20</td>
<td>45 minutes</td>
</tr>
<tr>
<td>TS3</td>
<td>Software consultancy</td>
<td>Business Development Manager Big Data</td>
<td>German</td>
<td>2017-03-23</td>
<td>65 minutes</td>
</tr>
<tr>
<td>TS4</td>
<td>Big Data analytics provider</td>
<td>Pre-Sales Manager for Industrial Big Data</td>
<td>English</td>
<td>2017-03-28</td>
<td>45 minutes</td>
</tr>
<tr>
<td>TS5</td>
<td>Software consultancy</td>
<td>Business Development Manager Big Data analytics</td>
<td>German</td>
<td>2017-03-29</td>
<td>35 minutes</td>
</tr>
<tr>
<td>TS6</td>
<td>Big Data analytics provider</td>
<td>Software Architect &amp; Marketing Manager</td>
<td>German</td>
<td>2017-03-31</td>
<td>65 minutes</td>
</tr>
<tr>
<td>TS7</td>
<td>IoT platform provider</td>
<td>Sales Manager Industrial IoT solutions</td>
<td>English</td>
<td>2017-03-31</td>
<td>30 minutes</td>
</tr>
<tr>
<td>TS8</td>
<td>Software consultancy</td>
<td>Principal Consultant Industrial Big Data</td>
<td>German</td>
<td>2017-04-03</td>
<td>60 minutes</td>
</tr>
<tr>
<td>TS9</td>
<td>Industry 4.0 research institute</td>
<td>Project Manager Working 4.0</td>
<td>German</td>
<td>2017-04-04</td>
<td>75 minutes</td>
</tr>
<tr>
<td>TS10</td>
<td>Big Data analytics provider</td>
<td>Director Product Management</td>
<td>English</td>
<td>2017-04-04</td>
<td>45 minutes</td>
</tr>
<tr>
<td>TS11</td>
<td>Software consultancy</td>
<td>IT Consultant for Big Data implementation</td>
<td>German</td>
<td>2017-04-05</td>
<td>60 minutes</td>
</tr>
<tr>
<td>TS12</td>
<td>Big Data analytics provider</td>
<td>Director Research &amp; Development</td>
<td>English</td>
<td>2017-04-05</td>
<td>30 minutes</td>
</tr>
<tr>
<td>TS13</td>
<td>Software provider</td>
<td>Manager Industry 4.0</td>
<td>German</td>
<td>2017-04-10</td>
<td>60 minutes</td>
</tr>
<tr>
<td>TS14</td>
<td>Big Data analytics provider</td>
<td>Head of Operations Industrial Big Data</td>
<td>German</td>
<td>2017-04-11</td>
<td>40 minutes</td>
</tr>
<tr>
<td>TS15</td>
<td>Software provider</td>
<td>IT Administrator</td>
<td>English</td>
<td>2017-04-11</td>
<td>45 minutes</td>
</tr>
<tr>
<td>TS16</td>
<td>Software provider and consultancy</td>
<td>Senior Consultant Big Data implementation</td>
<td>German</td>
<td>2017-04-11</td>
<td>50 minutes</td>
</tr>
<tr>
<td>TS17</td>
<td>Software provider</td>
<td>Director Product Management Big Data</td>
<td>German</td>
<td>2017-04-19</td>
<td>40 minutes</td>
</tr>
</tbody>
</table>
4 Methods
4.5 Critical evaluation

4.4.3 Manufacturing companies

From the manufacturing company perspective, a total of five qualitative, semi-structured interviews were conducted. Details can be found in Table 7 below.

Table 7: Interviewed persons at manufacturing companies

<table>
<thead>
<tr>
<th>ID</th>
<th>Company</th>
<th>Employees</th>
<th>Role</th>
<th>Language</th>
<th>Date</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC1</td>
<td>Adhesive manufacturer</td>
<td>1000 at interview site; 50000 globally</td>
<td>Big Data Scientist for production analytics</td>
<td>English</td>
<td>2017-03-21</td>
<td>45 minutes</td>
</tr>
<tr>
<td>MC2</td>
<td>Electronic manufacturing services</td>
<td>160 at interview site; 200 globally</td>
<td>IT Project Manager for Industry 4.0</td>
<td>German</td>
<td>2017-04-05</td>
<td>65 minutes</td>
</tr>
<tr>
<td>MC3</td>
<td>Commercial vehicle manufacturer</td>
<td>9000 at interview site; 33000 globally</td>
<td>IT Project Manager Big Data &amp; Cloud in production area</td>
<td>German</td>
<td>2017-04-10</td>
<td>35 minutes</td>
</tr>
<tr>
<td>MC4</td>
<td>Tobacco products manufacturer</td>
<td>5000 globally</td>
<td>IT Business Liaison Manager</td>
<td>English</td>
<td>2017-04-12</td>
<td>120 minutes</td>
</tr>
<tr>
<td>MC5</td>
<td>Bearing manufacturer</td>
<td>4200 at interview site; 48000 globally</td>
<td>Head of IoT Technology for production optimization</td>
<td>German</td>
<td>2017-04-19</td>
<td>65 minutes</td>
</tr>
</tbody>
</table>

4.4.4 Experts

In addition to the technology supplier and manufacturing company perspective, five experts for Big Data projects from a large management consultancy firm in Germany were interviewed for further input and additional information on a higher level to connect the dots.

Table 8: Interviewed persons at a management consultancy

<table>
<thead>
<tr>
<th>ID</th>
<th>Company</th>
<th>Role</th>
<th>Language</th>
<th>Date</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>EX1</td>
<td>Management consultancy</td>
<td>Senior Partner</td>
<td>German</td>
<td>2017-04-05</td>
<td>30 minutes</td>
</tr>
<tr>
<td>EX2</td>
<td>Management consultancy</td>
<td>Principal</td>
<td>German</td>
<td>2017-04-07</td>
<td>30 minutes</td>
</tr>
<tr>
<td>EX3</td>
<td>Management consultancy</td>
<td>Principal</td>
<td>German</td>
<td>2017-04-07</td>
<td>70 minutes</td>
</tr>
<tr>
<td>EX4</td>
<td>Management consultancy</td>
<td>Principal</td>
<td>German</td>
<td>2017-05-04</td>
<td>40 minutes</td>
</tr>
<tr>
<td>EX5</td>
<td>Management consultancy</td>
<td>Partner</td>
<td>German</td>
<td>2017-05-19</td>
<td>30 minutes</td>
</tr>
</tbody>
</table>

4.5 Critical evaluation

This chapter ends with a critical view on the proposed methods, since this needs to be considered when looking at this research’s findings (Collis & Hussey 2014). The critical evaluation is performed along the four criteria for validity and reliability assessment as proposed by Gibbert et al. (2008).

4.5.1 Internal validity

Internal validity refers to the degree of logical reasoning and causal links of the research to be “compelling enough to defend the research conclusion” (Gibbert et al. 2008, p. 1466). Hence, internal validity is closely linked to the degree of rigor during the data analysis phase (Yin 2009). To account for an adequate degree of internal validity, the following two measures have been applied. First, a research framework was developed as a basis for the formulation of interview questions as well as for deducting the codes for data analysis and showing clear relationships between challenges, resources and capabilities (Gibbert et al. 2008). Second, pattern matching was performed in the discussion chapter to compare this
research findings to other study’s results (Gibbert et al. 2008). Nevertheless, data triangulation was proposed as a third measure for internal validity, which could not be considered here, since only qualitative data from interviewees were collected.

4.5.2 Construct validity
Construct validity refers to “the extent to which a study investigates what it claims to investigate” (Gibbert et al. 2008, p. 1466). Construct validity is determined during the data collection phase and decreases when subjective judgements are used instead of objective measures to account for an “accurate observation of reality” (Gibbert et al. 2008, p. 1466; Yin 2009). One initiative taken to increase construct validity was to provide a high degree of transparency in this research by describing in detail how conclusions were made from the research findings and hereby build a clear chain of evidence (Yin 2009; Gibbert et al. 2008). For example, the methods how data have been collected and analysed were described in detail. Again, another potential measure would have been data triangulation, however this was not possible to achieve as mentioned above. Yet, in a few cases respondent validation was performed by sending analysed and interpreted interview transcripts back to the interviewee to allow re-checking the data and if necessary suggesting improvement ideas (Gibbert et al. 2008; Anderson 2010). These improvements let to refined and more objective research findings and thus contributed to the overall construct validity of the thesis.

4.5.3 External validity
External validity is often referred to as generalizability of the research and describes the generalization of the findings to a larger population (Gibbert et al. 2008; Collis & Hussey 2014). Although the case study research design performs poorly when it comes to generalizability, there are some measures to enhance external validity. By collecting data from multiple case studies and analysing these data across cases, generalizability can be increased (Eisenhardt 1989; Yin 2009). In addition, the reasons why these particular case studies were selected and some context details provide the rationale for the sampling (Gibbert et al. 2008). Regarding the degree of transferability to another setting, the geographical delimitation and industry focus limits it to German and Swedish manufacturing companies. However, it can be assumed that in every other country with similar characteristics in terms of competitive landscape, political climate, geography and demographics like Germany and Sweden the research findings can be applied too. Therefore, it is suggested that the reader decides for him/herself the degree of transferability of this research’s findings.

4.5.4 Reliability
Reliability measures the degree to which there are differences in the outcomes if the research was to be repeated (Gibbert et al. 2008). In other terms, a high reliability would mean that the research would arrive at the same results if repeated over and over again (Collis & Hussey 2014). Due to the nature of qualitative research, reliability cannot be considered as high, as the reproducibility will be low. Thus, it is likely that another research would not arrive to the same conclusions. As a countermeasure; the transparency of research procedures was improved by providing detailed information of how the entire research process was carried out (Gibbert et al. 2008). Yet, another limiting factor is that data analysis was conducted by one author only and hence did not involve other people during this process. Thus, a potential bias of the research needs to be considered here.
5 Results and analysis

5.1 Structure of results

In this chapter, the results from the acquired empirical material are presented. In more detail, the insights and learnings gained from the semi-structured interviews with technology suppliers, manufacturing companies and experts are described.

5.1 Structure of results

To serve as an orientation after reading the methods of data collection and analysis, the research problem and purpose are repeated to help the reader understanding the results better. The research problem was defined as follows:

“With the large potential of data variety waiting to be exploited, manufacturing companies are not sure which capabilities are needed to successfully transform towards a data-driven organization utilizing internal as well as external data sources”

Therefore, the purpose of this thesis was to investigate which resources manufacturing companies have to acquire and which capabilities must be developed to overcome current challenges related to the implementation of Big Data analytics by utilizing internal and external data sources.

The findings are structured along the data value chain framework by Miller & Mork (2013), which also serves as the process framework of the conceptual model in chapter 3.4. Furthermore, the structure also follows the logic of the research questions. After the results for each research question are outlined, an analysis follows. Since many interviewees did not want to be recorded on audio, only a few audio file transcripts were available to extract verbatim quotes to support the findings (Sandelowski 1994). Limiting quotes to only few interviewees from the overall sample would introduce a bias and further could threaten the anonymity of respondents. For these reasons, it was concluded to not use direct verbatim quotes. Nonetheless, to provide evidence and introduce some illustration to the findings, examples from practice – if applicable – were provided. This was done for two reasons: First, to provide evidence from practice that these challenges or topics exist and second, to provide an illustration of the theme for better understanding. These examples from practice are stories which were told by the interviewees as they encountered them in practice.

5.2 Data sources for Big Data

Data discovery concerns the identification of data sources, the purpose for data collection (i.e. identified business value) as well as getting access to these data sources and prepare data for subsequent data integration.

5.2.1 Data sources and types

According to three interviewees, there are two considerations, which have to be made when thinking about integration of data sources. Firstly, it must be determined which data can be found where. Secondly, data must have an identifiable business value, meaning that the integration of data sources should follow the clear purpose of adding value.

Almost all interviewees pointed out that potential data sources are basically present everywhere where data are generated and available. With regards to the potential areas of data collection, eight interviewees distinguished between internal and external data. Six interviewees indicated the integration of new data sources vertically and horizontally to the manufacturing company. Vertical integration relates to the acquisition and integration of new data within the company boundaries while horizontal integration focuses on capturing data.
5 Results and analysis
5.2 Data sources for Big Data

alongside the value chain both backwards towards the suppliers and forwards towards the customers.

Internal data and vertical integration
Internal data are described by interviewees as historical, structured data originating from the different IT systems the company currently has, such as the ERP system, MES or CRM systems. Those data can be financial data, production data, product-related data or organizational data related to certain departments, employees or processes (e.g. logistics). Most of the data, however, are machine generated data with the majority coming from the operations of machines on the production floor. Data sources on the shop floor are additional machine data such as log data, process data, quality data, maintenance data, production data, raw material data, protocols and sensors for heat, vibration, temperatures, pressure, cycle times, speed (rpm), energy consumption, sounds and noise, pictures and videos. However, not all machines on the shop floor can capture different data types. Older machinery and production assets do not have the software or connectivity features to collect operating data. In these cases, sensors can help to retrofit older production equipment or to collect additional data from newer machines. Further, sensors not only can be applied to machinery but also everywhere along the production process to collect additional process data. Which data types sensors can capture depends on their specified functions, whether they are audio-visual, optical or physical measurement sensors. These sensors are producing large data streams with high volume but only very little information. While some interviewees argue that these sensor data can only be captured in raw format, others state that these data flow in a structured format such as logs with a time stamp or location data. When looking across the shop floor boundaries, there are multiple other potential areas from which data could be sourced. According to interviewees, data from other departments such as human resources, finance, logistics, procurement or IT can be used to enhance data transparency for the production process. Another potential area of data collection does not necessarily need to come from machine-generated data but also human-generated data. For example, two interviewees pointed out that it is beneficial to digitize and analyse written maintenance or service reports to provide text analytics. These data are classified as unstructured data and further miss out on meta information. Further the quality of these data depends on the quality of the manual input.

External data and horizontal integration
In addition to internal data capture and vertical integration, six interviewees mentioned that manufacturing companies are looking for the capture of external data and horizontal integration of data as a next step. The interviewees emphasized that the utilization of external data is a newly observed trend, which supports manufacturing companies in enhancing their internal data base with externally acquired information. External data are useful to capture when the company wants to increase its data base by adding data which come from the environment, customer markets as well as from the supply chain in which the company participates. Four interviewees stated that environmental data are a potential source of external data. Hereby, data sources from the close environment of the company such as weather-related data (temperatures, air humidity or weather forecasts) can be collected. Those data can be of value if a company operates assets in remote, outside locations. Other data sources from the environment can be used for logistics optimization such as transportation data, traffic data or strikes. Manufacturing companies can also buy data from providers, which hold information about socio-demographics, micro- and macroeconomics, geographical data such as maps, weather or market trends and competitor behaviour. Besides sourcing external data from the environment, enhanced data capture along the supply chain of the manufacturing company is also considered along the axis of horizontal data integration. Data and information flows occur along the entire value chain, from raw material and components suppliers, incoming goods, manufacturing, logistics and customers. Although this area of data collection was regarded as new by two interviewees, five interviewees pointed out that it is not new and has been known long before
5 Results and analysis
5.2 Data sources for Big Data

but was until today very difficult to achieve. Now, with new technology emerging and better tools to enable secure data integration and exchange, it becomes possible. Moreover, horizontal integration can occur by integrating data sources towards the customers in order to connect with them and understand them better. Here it has to be distinguished between B2B and B2C data. There are large amounts of B2C data, which can be captured to understand the customer better such as social media data. This is a large potential waiting to be tapped, but there is also the question of customer privacy protection concerns.

Value of data
Further, as noted by six interviewees, there are four major fields where the discovery, integration and exploitation of many different data sources can create value:

1. **Process improvement**: Utilization of data sources to improve processes, such as optimizing manufacturing or coordination in the supply chain.
2. **Customer understanding**: Understanding the customer better by analysing social media data to get to know the customer’s beliefs, preferences and values. Used to provide products, which are mass-customized, cross-selling potential, individual targeted marketing or build individual products.
3. **Market analysis**: Analysis of social media data to understand what the consumers and customers thinking about the company’s products or the company in general.
4. **Product improvements**: Developing of intelligent products, often equipped with IoT sensors, to better facilitate understanding of the product use to subsequently improve it.

Although interviewees identified four applications where new data sources can create value, current data usage at the interviewed manufacturing companies is mostly directed towards the improvement of internal production processes. In the future, however, data usage related to customers’ understanding and product improvements is considered.

*Evidence from practice:* Interviewees from MC1, MC2, MC4 and MC5 focused on the first application field, namely to achieve process transparency in production for improvement. The manufacturing companies started with a review about their current process and which data they can additionally capture and then examined the data lacks and how to close them in order to move towards full process transparency. Only MC3 considered using data to improve their products, but this was in an early stage.

5.2.2 Purpose of acquiring new data sources

Overall three purposes of new data sources have been stated by most interviewees, which are presented below.

Firstly, new data sources are acquired to close current data lacks. Currently, machines on the shop floor generate large amounts of data, however each machine is only capturing its own operating data. Thus, data capture along a production process only occurs partially on the machines and not holistically along the whole process. To achieve full process transparency there is a need to transition from machine monitoring to process monitoring to be able to monitor the flow of products along the whole process.

*Evidence from practice:* As the interviewee from MC1 outlines, it is beneficial for data analysis to have all required data available, such as supply chain data, manufacturing data and quality data. If there are data gaps, then it must be determined if those data are needed. At least 90% of the data required should be available. The very first step was to increase the transparency of the current manufacturing process to determine which quality level of the finished adhesives will be achieved. At that point in time, the adhesive quality could only be determined afterwards and not prior to production. Thus, data transparency had to be
5 Results and analysis
5.2 Data sources for Big Data

increase to be able to explain why a certain quality standard had been produced. Therefore, the company implemented a new MES software, which can provide interfaces to different machines to collect all machine data in one place. Next, the company needed to close current data lacks, which occurred along the process. For example, they added new sensors to existing machines and captured data of the raw material from their suppliers.

Secondly, new data sources are not only explored and acquired to close current data lacks but also to gain additional information from data to either generate new insights or enhance current knowledge and understanding. According to one interviewee, the optimal way to go is to capture all data related to an identified phenomenon or problem, following the principle of ‘the more the better’. Only with a large base of different data it is possible to identify significance, correlations and determine causality between certain events, which is for instance important for root-cause analysis of machine or production failures. Thus, more data sources are beneficial as they provide new information. New data sources can also enhance knowledge. Although the shop floor workers have a lot of expertise when machine failures or operational disruptions are likely to occur, enhanced data capture gains additional insights, which can provide the certainty aspect to confirm the guess or expertise of the shop floor worker. It may not be surprising, but most information from new data is already known by shop floor workers. However, the true, new value lies in new discoveries in the data. How much these data are of value depends on the many different data and their relationship to each other. Thus, the value of data increases with volume and diversity of data. According to one interviewee, enhanced data capture can identify up to 90 correlations with a certain event and these correlations are of value as they were unknown before. Thus, the more data are captured, the more influences can be tested to discern which parameters are contributing to a certain phenomenon. For example, as one interviewee pointed out, an engine manufacturer was able to determine with the help of BDA that an open window during summertime caused cracks in the moulded engine blocks due to cold air flow coming from the open window. Therefore, it was valuable to increase data capture to explore different possible influences for root-cause analysis.

Evidence from practice: MC1 had to collect more data to predict upfront which product quality will be produced. They collected additional process parameters and raw material data from the supplier. With many data now available, they were able to conduct root-cause analyses to identify all possible factors influencing adhesive quality. This helped them eventually to redesign their production process and determine upfront which production parameters and raw material requirements were necessary to produce a batch of adhesives with desired quality standards. Since beforehand product quality could only be determined after production, the company saved inventory costs and decreased time to fulfil customer orders. MC2 achieved full data transparency to track product failures and subsequently perform root-cause analysis to see at which machine and operating parameters the mistakes occurred. They expect to detect quality problems in advance and therefore lower costs. MC4 captured additional process parameters to cope with tighter regulations regarding the quality and ingredients of its product. Data transparency along the entire manufacturing process helped them to fulfil regulatory obligations and deliver reports with all required parameter data.

Thirdly, new data sources along the supply chain (horizontal data integration) are acquired to strengthen business relationships due to shared incentives. For example, companies can benefit from an increased supply chain transparency in order to increase the traceability of parts. It is often necessary to proof what has been installed where and from whom. For now, the traceability works on the level of production cycles or batches, but perfect traceability requires to have this level of information for each single product, due to an increasing degree of customization towards customers. Further, data capture at the supplier side can be used to evaluate and negotiate with suppliers, for example when it comes to the quality of supplied materials. These are very sensitive data but for two reasons are important to know
5 Results and analysis
5.2 Data sources for Big Data

for the manufacturing company. Firstly, to use these data to adjust production processes for optimal production depending on the raw material quality and secondly to evaluate the supplied material from each supplier as a basis for contract negotiations. At last, increased data transparency would support automation of supply chains since data transparency benefits coordination between suppliers, optimized logistics and better management of risks.

Evidence from practice: As stated before, MC1 added new data sources not only along the production process but also performed horizontal data integration by adding raw material quality data from their suppliers. The interviewee stated that these data helped to adapt the production process and parameters of machines to the material quality. Another example was given by the interviewee of MC2, who stated that the company is considering acquiring additional data from its suppliers to increase the traceability of their manufactured products along the value chain. These data shall be provided to their customers to have an overview of the part history and further it will help MC2 and its suppliers to better understand why and where product failures occurred. MC4 has recently opened an e-commerce site, which is the first approach to get data from the end customers, so marketing and sales are eager to get data from the customers. They are currently exploring options to see how they can leverage these data for understanding the customer better and improve their products according to the customer’s preferences.

5.2.3 Analysis

When summarizing the findings regarding data sources from the interviews so far, the following statements can be extracted:

- Companies recognise the volume, variety and value of data
- Companies aim to combine internal and external data
- Interviewees identified the need for internal and external sources
- Interviewees indicated vertical and horizontal data exploration

When these four statements are contrasted to the reviewed literature, the following can be analysed. The proposed five Vs of Big Data (Laney 2001; Chen et al. 2014; Yin & Kaynak 2015; Gandomi & Haider 2015) have only been partially named by the interviewees. The three Vs related to the volume, variety and value of data can be found in the findings, however no interviewee indicated at this stage any thought or consideration on the other two Vs, those being velocity and veracity of data. Yet, on a positive note, interviewees confirmed that manufacturing companies seek new data sources as possible value opportunities, as stated by Addo-Tenkorang & Helo (2016). Therefore, it makes sense for manufacturing companies to increase the value of Big Data by striving for large data volumes and simultaneously for data variety to increase data value (Russom 2011; Chen et al. 2014). Since the interviewees outlined many different data types as possible data sources for BDA, the need for a large degree of data variety can be seen. Further, the interviewees also mentioned structured, semi-structured and unstructured data (Intel Corporation 2014). They also pointed out, where to find which kinds of these data in the same way as Gandomi & Haider (2015), indicating that interviewees are in general firm in knowledge when it comes to the characteristics of Big Data.

Regarding where to find possible data sources, almost all interviewees confirmed the classification of internal and external sources as proposed by Zhao et al. (2014). This is in line with the aforementioned trend to strive for data variety by utilizing both internal and external data sources (Zhao et al. 2014). Comparing the data sources identified by the interviewees with the five data sources mentioned by George et al. (2014), four out of five data sources are mentioned by both. Interviewees acknowledged internal data as private data, coming from business operations and transactions but also external data as private data, when they are provided by suppliers or customers. Further, public data were also
5 Results and analysis

5.3 Data discovery

named as a possible external data source. Even self-quantified data were named as a source of internal data. However, according to the interviewees there are, for now, no intentions to collect community data to capture social trends. At last, the data source classification structure by Hartmann et al. (2016) can be used to analyse the findings from the interviews. From the eight data sources the authors proposed, five data sources have also been stated by the interviewees. Those six are (1) already existing data in the IT systems, (2) self-generated data from sensors or employees, (3) data which can be acquired from data providers, (4) customer and supplier provided data and finally (5) freely available open data. Thus, according to Hartmann et al. (2016), crowd-sourced data, social-media data and web-crawled data are currently not considered by manufacturing companies. It seems like that manufacturing companies are well aware of not only focusing on internal data but also considering external data to complement internal data (Abbasi et al. 2016). Moreover, while internal data are limited to already existing data and new, self-generated data by sensors or employees (e.g. digitalized maintenance reports), there are more possibilities for potential data sources beyond the company boundaries. Since interviewees thought of the possibilities to collect data from both suppliers and customers as well as additionally acquiring data for a fee from data providers or connecting to publicly freely available data, it can be concluded that there is an adequate degree of large enough data volumes and data variety in both, data types and data sources.

5.3 Data discovery

Identified challenges by the interviewees related to data discovery can be further broken down into challenges related to the collection of data sources and challenges related to data access and data preparation.

5.3.1 Collection of data sources

In total, the interviewees listed challenges on all four levels, namely the technical, managerial, individual and organizational level.

Technical challenges

Current IT architecture

According to several interviewees, manufacturing companies already have databases from the proprietary IT systems such as ERP system, product management system, customer relationship system or MES for real-time production data. It is very common that companies have a dedicated IT system for every function, department or process with a local database. Data are stored in different databases in so-called data silos across the whole company, meaning each department has its own local data base. In today's manufacturing companies, often more than one data warehouse (usually 4-5) can be found, resulting in a variety of data silos with no connection between them for data transfer. According to four interviewees, this fragmented storage of company data from different departments poses a difficulty in determining which data are already available and which data need to be collected additionally. Each of them contains data but not in the same structure and frequently the same data sets are stored in more than one warehouse. Subsequently, data quality problems are caused by this unorganized data management and fragmented IT infrastructure.

Evidence from practice: MC1 had two main data warehouses, one each for material quality data and manufacturing data. Further, MC1 had several manufacturing sites where each site locally collected and stored data. Thus, MC1 faced the challenge to integrate data stored in many different warehouses in one central database. The interviewee of MC3 illustrated their IT infrastructure as fragmented: They have one ERP system from Microsoft, which takes
5 Results and analysis
5.3 Data discovery

care of transactions related to purchasing and sales. On top of the ERP systems operates a recently installed new system, which is used for production planning and set up times to support planning activities. On the shop floor, there are numerous IT systems, too. For example, they have different manufacturing software from each machine provider and one system to collect all machine failures, which operates on top of the MES.

Managerial challenges
Regarding managerial challenges, three challenges were identified by interviewees: Understanding the need for change, defining a data strategy and vision and financial investment.

Understanding the need for change
The need for change is arising from two directions: external market pressure as competitors begin to utilize Big Data techniques and internal motivational aspects as certain employees start initiatives towards Big Data. In most cases, the external market pressure triggers the need for change, as the competition intensifies. CEOs and top management must realize that digital transformation and new technologies such as Big Data will affect their business and that they must adapt rather sooner than later by thinking about the future and its challenges against the status quo. Furthermore, top management has to be aware of the potentials of digital disruption and need to place increased focus on this. There is a need to understand its opportunities (to improve and become better) and the risks of not doing it (competition will become stronger). To do so, top management of manufacturing companies should evaluate the potential of Big Data and Industry 4.0 for their company. For now, these concepts are regarded as buzzwords with different meanings to different people. Since these buzzwords are used as a marketing instrument with an increasing emphasis on success stories, the expectations can easily be much higher as the reality as pointed out by five interviewees.

Evidence from practice: MC4 was driven by competitive pressure to innovate, as they feared to be put out of business if they do not utilize Big Data as their competitors. However, they did not rush in their decision, rather they informed themselves about the potential benefits and risks and then identified their needs and business value. They did not want to be first, but to be innovative for themselves. The interviewee of MC5 stated that the concept of industry 4.0 should be always questioned critically as it is not possible to apply the concept of industry 4.0 and Big Data in every manufacturing company. Further, companies should also question constantly how they can be disrupted and what they can do about it?

Defining a data strategy and vision
Once the need for change is understood, CEOs and top management have to establish a vision and data strategy to formulate and identify a clear business value, which was the most cited challenge by interviewees. Collecting data is an easy task, the more important questions is what to do with them. It is easy to get overwhelmed by the possibilities of a data eco system where no one can sense the opportunities for seizing. Next, top management has to ask themselves, which data should be regarded as a commodity and which data are critical and need to be protected and secured? Hereby companies could also think about those data, which can be of importance to achieve competitive advantage and are of high value. This can be considered as the unique selling point the company has compared to its competitors. Relying on increased data-driven decision-making requires trust in data and their information. Since the management also relies on data intelligence, it is vital for them to see the value in data and trust the outcomes of BDA.

Evidence from practice: The interviewee of MC3 argued that there is need for a clear data strategy in manufacturing companies. Just collecting massive amounts of data without any clear goal or aim does not make a difference. Moreover, the acquisition of external data does not provide a competitive advantage, since those data can be acquired from anyone.
5 Results and analysis

5.3 Data discovery

else, too. Thus, companies need to think about acquiring unique data sources and build unique data sets. At last, a data strategy also includes considerations about which data shall be shared with others to gain shared benefits and not risk business damages by sharing sensitive data. As the interviewee of MC4 further outlined, as a manufacturing company they work a lot with innovation. However, as the interviewee argues, being innovative does not indicate to be among the first to innovate. Instead, one must firstly understand what Big Data is and then transform it into business needs and business value. The interviewee of MC5 pointed out that they asked themselves the following critical questions, when they thought about a possible data strategy: How do I create value? Which processes do I need to improve? Where can I still optimize? How can I change my business model? Further, companies should start to think early on about new business models based on data, because this will be the future to not sell products anymore but to sell data-based services in connection to the hardware.

Financial investment

Another challenge related to the managerial level are financial investments. Those might be large, depending on the ambitions of the company and the status quo. The business case differs for each company, so do goals with the digitization, for example how to generate profit. The budgets are more linked to keep it running rather than providing data. There needs to be a mind-set of change that data has value, and this process is ongoing but it is not there yet. The financial situation is also tough. Further, as interviewees stated, company executives would rather invest in tangible assets such as machines than intangible such as software or IT infrastructure, especially if the return on investment is questionable or yet to be determined. It is difficult to justify an investment that is planned to be fully used in three years, since there is no direct value today.

*Evidence from practice:* The interviewee from MC4 observed the following: The general rule is that there are no investments to be made without a business case. The problem with new technologies such as Big Data is that there is no business case due to difficulties of determining the value upfront. This is difficult to understand for top management reaching ages of 50 years and up that IT is the new focus of investment.

Individual Challenges

*Appreciation of data value*

Just because data are available, does not mean that all data are relevant and should therefore be captured. It remains challenging to determine which data are needed, useful or even critical to acquire to better understand certain processes. Which data are needed depends on the use case and requires expert knowledge regarding the process parameters. For example, the manufacturing process operator is the one with the expertise and knowledge, who needs to decide about which additional data for a certain use case are currently missing and need to be collected. Interviewees emphasized that employees, with special regards to blue collar workers on the shop floor do not have a feeling for the value lying in Big Data. They prefer to rely on their own judgement and experience and therefore do not see the value in data showing the same observation or knowledge they knew before. According to one interviewee, around 80 per cent of efforts are needed to identify the right data and 20 per cent to actually collect data.

*Evidence from practice:* The interviewee of MC1, a Big Data scientist, said that it required the expertise from both data scientists and production workers to determine which additional data could be used to gain new insights. Hereby, the production workers contributed with their expertise regarding the physical process and the data scientists were responsible for the digital data. Then, they combined their knowledge to understand the data in relation to the process. The interviewee of MC2 stated that it was difficult to teach employees the value of data, what data means and how they need to change their attitude towards data capture.
5 Results and analysis

5.3 Data discovery

Also, they needed to be taught about data quality. The interviewee of MC4 pointed out that their current supply chain is too simple and therefore collecting sales or supplier data would not have an impact on bottom line improvement. Hence, they see no business case for now. However, they claimed that it would be interesting to know more about their supply chain.

Organizational challenges

Openness to change
Four interviewees addressed the challenge of achieving openness to change among employees. Interviewees revealed that successful conquering of challenges can only happen when all employees understand the need for change and are open to change. As one interviewee explained further, employees in Germany may be reluctant to change due to their low risk aversion. Besides, as another interviewee argued, the introduction of new technologies will also lead to job losses, since technology can take over or eliminate many jobs. For example, employees assist in implementing Big Data on the shop floor and contributed with their process knowledge and expertise only to build a data model and automate the process driven by Big Data to make their job obsolete in the medium-term. Thus, employees refuse to change and adopt new technologies such as Big Data. In contrast, a third interviewee argued that the adoption of new technology will help firms to stay competitive in the long run and therefore will secure or even add new jobs. Therefore, employees should be open to change since change will come along with benefits.

Evidence from practice: The interviewee of MC3 explained that since he got appointed as project manager to establish a Big Data platform in the company, he only deals with people issues, to establish a data-centralized corporate culture. There is a high refusal rate of Big Data among the employees and he needs to manage how an individual deals with the topic of Big Data as well as how Big Data come together with people. For example, as he points out, data are machine generated and provide rational and structured information, while people do not react like a machine, since people are also emotional. Thus, combining Big Data and people is very difficult at MC3.

5.3.4 Data access and preparation

Challenges related to data access and data preparation occur on the managerial, technical and organizational level.

Managerial challenges

Ensure and negotiate data access
How to legally obtain access to data or be allowed to collect data from a data source is of political nature. There is no general rule for getting access to data sources, as it has to be proven for each case in itself. Besides, data ownership and data privacy concerns are not easy to resolve and pose a challenge in both vertical and horizontal data integration. Open data from third party sources are relatively simple to access since paying a one-time fee or monthly subscription fee provides access to a customized report. With regards to horizontal data integration, one has to ensure to get access to the required data. Here it is beneficial to have a strong negotiation power or deep and long-term relationships with suppliers to facilitate data access. Two interviewees pointed out that the key to success for data access is to create mutual benefits of data exchange for both sides. For example, lock-in effects and the forming of long lasting relationships if data access and exchange is granted. The goal is to establish a supplier quality network where data transparency is key.

Evidence from practice: As stated before, MC1 also integrated data about the raw material quality from its suppliers during the Big Data project. Hereby, it turned out to be beneficial for the plant manager to have strong and well-maintained supplier relationships as this helped
5 Results and analysis

5.3 Data discovery

to a great extent to get access to these data sources from the suppliers. The interviewee of MC3 explained that he encountered several barriers while trying to obtain access to data from other departments, for example human resources or finance. The managers from the other departments raised concerns about data security and privacy, and the future use of these data, since in the early stages of the Big Data project it is impossible to foresee future data usage. For these reasons, most managers refused to grant access to their data. As a result, the project manager needed to get the support from top management to resolve these issues and get access to the data.

Technical challenges

Once access to a certain data source – internally or externally – is established the next step is to ensure technical connectivity in order to transfer or exchange data.

Internal data access through IT and OT integration

To get data access to internal data on the shop floor, the connection of production equipment, such as machines and belt conveyors to IT systems becomes necessary. The aim is to have the operational technology (OT) and IT combined and integrated as opposed to operate separately. Only if IT and OT integration is achieved, it is possible to collect and store all production and machine data along the manufacturing process. Challenges are related to the connectivity of machines and production assets to IT systems. Connectivity is an important enabler for integrating machine and IoT sensor data into IT systems. Thus, the challenge is to build a connectivity layer into the IT system to enable communication between IT and OT. However, multiple technical standards and connectivity formats increase complexity and preventing the built-up of such a connectivity layer. For now, there is a very immature market with multiple machine and connectivity standards. Which standards eventually become dominant is yet to be determined, thus manufacturing company need to decide wisely and with a bit of luck. Regarding internal data, which come from other departments such as human resources or finance, interviewees stated that getting access to these data is not technically challenging, as data can be accesses through internal IT systems or exchanged via the IT network.

Evidence from practice: The interviewee of MC2 explained that how much of production machine data can be captured depends on two factors: the manufacturer and age of the machine. Their manufacturing process consists of different five machines and all of them were built by different manufacturers. Each machine therefore has its own interface, different software and different data capturing standards. Further, the different machines on the shop floor can be of varying age. Older machines do not have the necessary interfaces and software to connect with other machines or IT systems. In contrast, modern machines have an integrated software interface, where machine-related data are collected, stored and integrated into other databases. Despite facing these challenges, MC2 aimed to reduce quality problems and detect product failures or anomalies along the production process, which required, data. As a first step, they implemented a new MES system, which can connect to all different machines along the production process. This was already a time-consuming and expensive project to achieve. Next, they needed to add additional sensors, mostly optical sensors and equipped each electronics part with an ID. Now, each electronic part can be traced along the process.

At MC3 the interviewee stated that data from the production process and machines could only be captured and stored, if the machine provides a connectivity interface. They have relatively old machines at one manufacturing process step, which are not able to connect to the ERP system and therefore no data can be captured. MC3 followed the approach to integrate IT and OT by exchanging physical equipment at the same time when they changed parts of their IT systems. They envisioned a step-by step plan for the next 3-5 years to upgrade the whole production plant. Hereby they always ensured zero operational
5 Results and analysis
5.3 Data discovery

Disruptions, which required systems to run in parallel until they were sure the new system works reliable.

The interviewee of MC4 explained that with regards to missing connectivity standards, they look at institutions and research facilities to monitor closely which standards they propose, use and prefer. Further, they try to adapt the same standards as large manufacturing companies to increase the likelihood of adopting those standards, which will become dominant in the future.

At MC5 the interviewee explained that their older machines need to be equipped with standardized gateways that enable data transfer from the machine to IT systems. This was their first step to collect data from production assets. The next step was to determine how data can be transferred from the machine gateway to the IT system. They evaluated several options, such as wired transfer via LAN or wireless transfer via Wi-Fi or 5G mobility network. In the end, the company chose a wired LAN transfer, as LAN was partially already available and required the least investment. However, they would also like to look at 5G connectivity to be prepared for the future.

External data access

For transferring data from external sources, two points must be considered: first, to figure out how to technically connect the data source with the system and second, to be very specific about data access credentials and data security. In most cases, there is already some data pipeline established with key suppliers, for example as part of an ERP system integration. However, as several interviewees pointed out, horizontal ERP integration is in a very early stage. There is also the possibility to transfer data via an agent, who acts as an intermediate and does not require the direct integration of ERP systems. When sourcing data from external sources, companies need to think about data security and data privacy – the data must be safe and clean. Other options are so-called data clearing services, which control all data leaving and entering the firm. By doing so, data can be filtered along selective criteria. At last, the transfer of data should follow common standards of communication exchange and data formats. A steady and secure data exchange is key for successful data transparency along the supply chain.

Evidence from practice: The interviewee of MC2 explained that the company wanted to automate their orders from their main customers by reading the incoming order document automatically and initiating the order fulfilment process. Since each customer has its specific order document, they tried to convince them to switch to one standardized order document, which the main customers did not want to do. Therefore, they contracted a third-party provider to transform the order documents in the desired structure to automatically analyse them using text analytics. MC3 preferred one central gateway for all data, which are exchanged with suppliers, as it is easier to monitor and control only one data pipeline in order to ensure high data privacy and data security. The more data exchange gateways they have the higher the security risk.

Organizational challenges

Two organizational challenges were identified by the interviewees in firstly being cross-functional coordination and cooperation and secondly the preparation and organizing of data.

Cross-functional coordination and cooperation

Another challenge confirmed by most the interviewees is the increased need for deepened cross-functional or cross-departmental coordination and cooperation in the era of Big Data. In vertical data integration, data originate from different sources within the company and connect processes and people over data. Therefore, employees also need to cooperate and coordinate among departments just like data are flowing. The IT department and the
5 Results and analysis
5.3 Data discovery

production floor in particular need to collaborate in Big Data projects. For example, when the BDA department intends to integrate more data sources from the production process, the IT department and the shop floor workers need to collaborate to capture more data. Another example was provided by two interviewees independently: Both highlighted the need for close coordination in purchasing of new equipment, which supports the data capture and connectivity standards as these decisions cannot be made in isolation anymore. In vertical data integration, getting access to other department’s data is not easy. However, as five interviewees indicate, inadequate intra-company communication hinders this. Challenges are the lack of coordination and communication between certain departments. For example, marketing and sales are not communication adequately with the production floor, nor they are handing their data easily over due to data concerns and mistrust. Thus, it is important to coordinate over processes and work together across departments to achieve the desired transparency of data.

Evidence from practice: MC1 was in the middle of an organizational change forced by the implementation of a Big Data project. A new organizational layout was necessary for two reasons: First, to facilitate communication and coordination between the IT department and production department as well as top management and second, to introduce new functions and roles related to Big Data management, especially data scientists. According to the interviewee of MC1, the company solved this challenge by creating a new role called “Business liaison manager”, which established the link between the production department, the IT department as well the business management. The tasks were to facilitate communication between all departments as well as to understand the needs of each department and how the other departments can cooperate to fulfil these needs. The key for success of this role is to communicate on a high level with each department and enable coordination and cooperation instead of internal competition and mistrust.

Preparation and organizing of data
Already available data at manufacturing companies are often not of high quality, as they are not organized in a structured way or sometimes are even partially missing, which is related to the fragmented IT infrastructure, many databases and data silo structure. This leads also to low data quality and data consistency. However, in the Big Data era it is required that data are clean, structured and stored in an organized way without data duplicity or redundancy. The challenging part is to organize and structure all data in one place, for example a large database, where the data can be accessed and used by several employees from the whole organization. According to interviewees, data labels for each type or source of data enable a structured and organized approach to store data. Thus, data definitions and data labels enable the multi-use of data sets. Data labels should be explained in a catalogue for everyone easily to be understood. Clear data labels and definitions must be present, otherwise the whole data structure becomes unorganized and difficult to expand on. Therefore, three interviewees stated that it is important to have a clear data architecture in mind in the beginning of a Big Data project. However, to determine which data structure or organization is ideal is difficult in the beginning. It is possible that during the progress of Big Data initiatives more and more data sources are added that ultimately require a modification to the current data structure and labels. Hence, hindsight, foresight planning and a bit of luck are beneficial.

Evidence from practice: MC3 intended to integrate data from sales and claims reports, which both originated from different systems. It required a lot of manual input in Excel to create one report showing data from both sources. They needed experts from both departments who could explain the data meaning since clear data labels were missing. They learned from this problem that it would be key to have easy to understand data labels in the future if they progress with integrating all data together in one database.
5 Results and analysis
5.4 Data integration

In the following the challenges related to data integration are presented. According to the interviewees, challenges arise on the technical, managerial and individual level.

Technical challenges
Technical challenges during data integration were twofold: At first, interviewees pointed out the challenge of building a powerful Big Data platform, which must be capable of integrating data from many different sources. At second, the to be integrated data need to be of a specific structure.

Building a Big Data platform
Interviewees revealed that the current IT infrastructure of manufacturing companies limits to a certain extent the ability of BDA. Overall the IT landscape in manufacturing firms can be described as fragmented and heterogenic. Due to stepwise or incremental upgrades as well as mergers and acquisitions the IT landscape is diverse. In large companies over 1,000 of such IT systems can be found. On another note, the IT infrastructure is often older than expected by outsiders and consist not only on purchased off-the-shelf solutions but also customized and in-house built solutions. There is need for new, powerful Big Data platforms that can not only integrate data from a variety of sources but also manage all data in one place. It is an obstacle to choose the appropriate platform technology, since there are numerous IT providers and the companies simply lack the expertise to evaluate which technological IT platform suits their needs best. Considering the fact that most likely a company starts with a single use case for Big Data and then gradually adds new ones, there is risk that the IT infrastructure becomes fragmented again as each use case requires different technology components (hard and software). Moreover, the replacement of old IT systems with a new Big Data platform poses several implementation challenges, since operational disruptions due to malfunctioning IT would have high impact on business operations. Further, companies struggle to decide which data storage requirements are needed. Current data warehouses cannot efficiently handle unstructured data that is why the data warehouse can be seen as storage for structured data. At last, companies must decide where to store the data. Data storage can occur on premises with servers or in a cloud, which can either be local or global (public, private).

Evidence from practice: MC4 had a lot of data in different IT systems stored on the shop floor but these systems are not interconnected. Data from the manufacturing process are stored in at least three different warehouses. When new IT systems are added or older ones get upgraded, the whole IT landscape becomes even more complex. Since analysing data stored in different warehouses is a difficult task and this is further limited by the fragmented IT infrastructure, MC4 concluded to change their whole IT infrastructure to be prepared for the Big Data era. They identified the need to replace all old IT systems by only one sophisticated system. In the end, they envisioned a single system enabling easy access to all data with the required flexibility. But this process will take time and needs to be planned in steps since the production has to continue running and operational disruptions would have a high impact. The interviewee outlined that this stepwise implementation is still a risky approach, but risks can be reduced by not changing the whole system at the same time. Likewise, MC5 stated that their first step was to clean their current IT infrastructure and simplify their data warehouses and IT systems. However, the interviewee indicated that it will not be possible to have a homogenous IT landscape, where all data are stored in one place. For example, the company operates several production lines in different countries and therefore the local IT will always be different.

Data structure
Key for successful data integration is to enable a seamless fit of new data to the current structure of the database. Thus, it is important to have the data organized in the same
structure as the stored data to allow a smooth integration. For this purpose, the interviewees advised before building a data framework upfront to determine the structure and organization of data (see *Organizational challenges of Data access and preparation*). Data source integration thus is complex and depends on how well organized the data from the data source are. This indicates that defined data architectures and structures also need to be pushed back to the data source where the data come from to receive the data in the preferred format. In order to have a solid structure of data, it is of importance to have clear data definitions, which must be shared to understand data meaning and content. The data can be integrated based on the time label, location label or measurement or common names such as ID numbers. Thus, it is important to coordinate all integrated data with the same data labels to have clarity what exactly each data set means. Moreover, aspects of data quality, such as data completeness, data accuracy and data definitions or duplications are important to consider. Otherwise the quality of data might decrease the overall value of the data and needs manual refinement.

*Evidence from practice:* The interviewee of MC5 explained the integration of machine data. Since the structure of machine data is varying over age and different standards of machines. However, all data have a time stamp by means of which they, in theory, can be linked together. In practice, it was revealed that data synchronization within milliseconds was required, which was not supported by older machines as they were collecting data only each second. Thus, it was more difficult to control the production in machine in real-time.

**Managerial challenges**

*Finding an implementation partner*  
Related to the investment and financing, there is a challenge to decide whether to implement Big Data alone or receive help from a partner (such as a IT consultancy firm). While only one interviewee encouraged manufacturing companies to do it by themselves to experience the highest learning experience and skill building, six interviewees talked about finding an implementation partner. There are three reasons for this: (1) manufacturing companies are financially restrained and therefore cannot fail often, (2) they do not have the expertise and skills to carry out such projects related to technology evaluation and implementation and (3) they do not know how to scale one use case of Big Data to the whole organization. However, the choice of a suitable and trustworthy partner is not a simple. The partner should evaluate technology and business needs tailored to the company and should also engage in a long-term relationship.

*Evidence from practice:* Due to the identified need of replacing all IT systems with one sophisticated Big Data platform, MC4 was looking for a potential implementation partner. MC4 wants to have a partner, to show them what can be potentially done by utilizing Big Data but also listens to them and understand their specific needs. Further, they would prefer only one vendor, which should be able to provide a holistic solution. The partner should support them along the implementation process with expertise in Big Data and IT topics, where MC4 does not have sufficient understanding.

**Individual challenges**

*Data security and privacy*  
Another aspect of a data-driven mind set relates to overcoming concerns with data privacy and security when it comes to the storage, integration and usage of personal data. Three interviewees argued that older generations are more concerned with data privacy and protecting their own data in contrast to younger generations. These concerns increase when the storage of employee-related data is necessary (for example, capturing personal data of machine operators and linking them to output levels). The interviewees concluded that data privacy and security would highly affect the acceptance of Big Data technologies. With new
5 Results and analysis
5.5 Data exploitation

regulations regarding data protection to be launched, these challenges will get even harder to solve.

Evidence from practice: The interviewee from MC3 outlined that in the company there is much distrust in data. He is currently in the process of building a centralized data warehouse for the whole company but faces challenges related to data protection and security as well as data privacy: Some departments refuse to supply their data since the future use of these data cannot be determined for now and remains vague. Further, the worker’s council forbids the storage of any personal data to protect employees. Employees also refused to make use of data storage in the cloud as they believe their data are more secure on servers within the company and should not be stored elsewhere. The interviewee argued that most employees do not know what a cloud is and therefore remain resistant to this storage option. He concludes that all these challenges related to data privacy, protection and security cannot be solved in the near-term.

5.5 Data exploitation

At last, the challenges related to data exploitation, the third and last step in the data value chain framework, are presented. Again interviewees named challenges on all four levels.

Technical challenges

Data analytics and visualization

There is no or only limited further usage of these stored data and the majority of data is not further processed and analysed and remain non-utilized. Until now, manufacturing companies only utilized a small fraction of their stored historical data in order to create reports for BI to provide the top management with an overview of business performance. The creation of BI reports consists of many manual steps whereby the data needed are collected from many different data sources and are re-organized and structured in Excel spread sheets to build tables and diagrams for the report. Moreover, data quality is an issue and affects the outcome of the report in a negative way, since data are drawn together from different databases and are processes manually. Therefore, interviewees argued that new IT systems must be capable of automated data analysis and visualization in order to exploit the information richness and value of collected data and use the analysis results to support decision making. The visualization of data supports employees to understand the insights the data revealed better and help them to act accordingly.

Evidence from practice: The interviewee from MC3 indicated that the BI department draws its data from numerous IT systems such as ERP systems, MES or department-specific systems. Since all data are stored in different systems, the BI department manually needs to create reports with Excel. Since this process is too time-consuming and due to the manual work steps is subject to mistakes and false numbers, the company strives for a Big Data platform with high data analysis capability and automated visualization of results. MC1 was already able to exploit new insights from data: With many data now available, they were able to conduct root-cause analyses to identify all possible factors which influence product quality of the adhesives. This helped them eventually to redesign their production process and determine upfront which production parameters and raw material requirements were necessary to produce a batch of adhesives with desired quality standards.

Managerial challenges

Facilitating cultural change

As according to interviewees, challenges related to change on the individual and organizational level need to be facilitated by top management. Thus, it is the top
management’s task to establish and environment to support and facilitate cultural change in the organization. Understanding the value of data, overcoming concerns of data privacy and security as well as achieve trust in data and a changing way of decision-making requires the whole organization to change towards a data-driven culture. Soft aspects such as open communication and a high degree of employee participation as well as addressing the risks and fears of employees are supporting the facilitation of corporate culture change. It requires a lot of time, openness and corporate trust and it is important to set small targets in the beginning. A step-by-step approach is advised and it is most importantly to bring all stakeholders to the table. This being said, there is no clear or direct path and therefore it needs open communication and coordination to be successful across all hierarchies and departments. There should be no direct result or target envisioned but the target should be left open to see where the company is heading and take from there the next step. According to one interviewee, one example is that certain employees dedicate 10 per cent of their time to learn about Big Data to present the ideas to colleagues and build up expertise. Thus, trainings and education of employees are important that everyone understand the need to change.

Evidence from practice: The interviewee of MC3 suggested the following: During implementation, talk to the people on the shop floor, resolve the issue of trust in the data and motivate them. Moreover, a good knowledge management is needed, open communication culture, new skills and knowledge development and to let the employees decide and actively participate. Also, learning is key and should not be neglected. The company MC4 saw themselves confronted with a diverse workforce. Their employees ranged from 18 years to over 60 years. This required different education methods, as older people are not that IT affiliated compared to the younger generations. The interviewee of company MC4 argued that organizational change management cannot only be driven by the IT department, a holistic approach has to be adopted. Only if all departments together work on a holistic solution, change can happen. At last, the interviewee of MC5 revealed the following: The older the employees are, the more reluctant they are when it comes to change. However, he is confident that older employees will adopt too, it takes just much more time.

Individual challenges
Interviewees outlined two individual challenges related to data exploitation. Firstly, individuals need to acquire new skills and knowledge to exploit the value of data and secondly, individuals have to trust the insights which are revealed by data.

New skills required
All interviewees were consistent in their answer that the emergence of Big Data leads to a changing skill set towards more digital or IT-related skills. In the past, the skill set for blue collar workers on the shop floor consisted of mechanical engineering competencies and substantial expertise related to the production process and running of machines. The roll-out and implementation of Big Data projects across all company operations requires expertise in both IT and process knowledge. Employees need to acquire the ability to transform Big Data into Smart Data. Data in itself do not provide any insights and are therefore not of value. Information and value of data are only revealed when data is put in relation to other data. Data are showing a picture of the digital world related to a physical process. Therefore, employees have to be capable of transferring data insights from the digital world to the physical environment. Yet, it is difficult to incorporate both the data and process knowledge. Thus, the current skill set needs to be adjusted to new technologies. While the majority of interviewees argued that employees need to extend their current skill set by acquiring these new IT and data skills, few interviewees were of the opinion that IT and data skills will replace mechanical engineering competencies over time. One interviewee highlighted that the future requires generalist workers having a rather broad skill set related to process knowledge and digital technologies rather than being deeply specialized in one area specifically.
5 Results and analysis
5.5 Data exploitation

Evidence from practice: The interviewee from MC4 reflected on the need for new roles and skills: The company saw the need for data scientists and so-called “data cowboys” who possess unique skills to identify the value in data and put large data sets together. In the production their workers provided the necessary process knowledge and mechanical background but lacked data understanding. Therefore, data cowboys and production workers were asked to work together to understand the data in relation to the process. The interviewee further expressed concerns about a possible talent shortage in the future. He expressed the importance of hiring high potentials from outside the company who will bring fresh knowledge, skills and views into the company. He also said that companies need to educate their own employees to learn the skills, since every company will have their own specific way of implementing Big Data.

Trust in data insights
Closely connected to appreciating data value and overcoming data security and privacy concerns is the building of trust in data. Several interviewees explained that in nearly every case when first results of Big Data projects were shown to employees they refused to trust these data. This was especially true when the data showed results contradicting to the employees’ own views or beliefs. However, as two interviewees mentioned, it is important to trust the data for the following reason: Current IT systems can display BDA but in the end humans have to decide whether the analysis results are trustworthy and can be used as a basis for data-driven decision making. This requires not only trust in data, but also the ability to judge whether these data represent a phenomenon or problem of the production process in a reasonable way.

Evidence from practice: At MC1 data trust issues were encountered: When BDA was used to analyse the current performance of the process and the results were presented to the production workers, they refused the analysis results, since they did not want to trusted the data. Obviously, the data showed several inefficiencies across the process yet the production workers did not believe it. The employees were not able to accept the findings, since they worked on the plant for over 20 years and rather relied on their own experience and intuition. Eventually, it took them almost one year to accept the results from data analysis and trust the results. Similarly, the interviewee of MC4 stated that the company was facing the challenge to accept trust in the data analysis results. He was confident that sooner or later the employees will accept, but in his opinion this step is crucial to achieve.

Organizational challenges

Changing work processes
The change of work processes is impacted by three aspects: First, Big Data technology requires different working environments, second, decision-making will be based on data and not on experience and intuition and third, Big Data demands additional work tasks to be carried out. At first, the way of working is changed due to the emergence of Big Data. Three interviewees stated the fact that for Big Data to be successful a shift from the traditional working culture towards a more entrepreneurial culture is required. For example, working with data analysis for BI purposes is fundamentally different than working with BDA using data from many sources. BDA demand new analytical methods and technological concepts and therefore poses a major shift from traditional BI processes which were performed the same way for the last 20 years. Thus, the BI department and the new department responsible for BDA have different working cultures and therefore have difficulties working and acting together as one entity. The second aspect is related to the change in decision-making. One interviewee explained this by means of the following example: Prior to the usage of BDA to improve the production process or predict machine failure, shop floor workers could detect and solve operational problems based on their knowledge and expertise. This expertise was built over several years of working. With BDA, decisions are
5 Results and analysis

5.6 Summary of results

not based on experience and intuition anymore, but data-driven decision-making relies on rational and neutral data algorithms and therefore changes the decision-making process in the future. This aspect is also closely related to the individual challenge of trusting data insights. The third aspect is linked to additional tasks that need to be carried out with Big Data technology. Until now, the production workers were focused on achieving operational goals and output volumes and prevent operational disruptions. During data exploitation, new tasks will be added to the daily working tasks of employees, relating to identifying new use cases to apply data analysis results to other problems or improvement initiatives. The fundamental issue here is that those are different interests and motivations: The production workers need to achieve operational objectives and are therefore not concerned nor do they have the resources to think about new use cases or to collect even more data to enhance data analysis results.

**Evidence from practice:** As the interviewee from MC4 outlined, the BI department is the intermediator between IT and controlling department. BI is very structured, but now BDA is more open to new results and new insights and more related to processes. Their new BDA department is different, like a start-up with an entrepreneurial culture. The interviewee of MC3 points out that the organizational boundaries pose numerous challenges to overcome when implementing a Big Data project. There are two different ways to start a Big Data project: Either the Big Data project gets its own analytical department separated from the traditional BI department or both are integrated together. Nevertheless, as he argues, in both cases the mentality and culture of two worlds are crashing together. Moreover, according to the interviewee from MC1, they were thinking about retrofitting older machines with sensors to capture additional data if these data can provide more information to current data analyses. However, considering high staff shortage, the people on the production floor were already overloaded with fulfilling operational goals and therefore did not have time to think about additional data capture. The interviewee of MC1 also outlined at this occasion that enhanced data visualization tools would help production workers to process information faster and therefore could potentially reduce their workload.

**Cultural change**

Almost all interviewees agreed that the organization needs to undergo a change of corporate culture to benefit from Big Data by creating a data-driven culture. As several interviewees recognized, changing the culture will be a slow process due to the traditional and risk-aversive working culture in the manufacturing sector. According to the interviewees, the success of cultural change depends on the following three factors: Firstly, the speed of new technology introduction in the workplace and the subsequent organizational roll-out. Secondly, soft aspects of change related to communicating the overall direction and need for change from the top management. Thirdly, the first experiences made with Big Data technologies. Hereby, early successes are enabling change while failures and drawbacks as well as negative experiences are hindering change. As two interviewees stated, there is a barrier to start a first Big Data initiative and therefore its success or failure determines further progress or refusal. At last, the success of cultural change also depends on the desired targets and what is envisioned to be achieved. Concluding, the interviewees argued that cultural change is only possible and successful when there is a clear benefit of change communicated and experienced by the organization.

5.6 Summary of results

The findings from the interviews highlight several challenges manufacturing companies face on the technical, managerial, organizational and individual level when moving along the certain steps in the data value chain as presented by (Miller & Mork 2013). In total, the interviewees identified 21 challenges. Even though some challenges seem large, the interviewees also partially proposed some solutions in terms of to be acquired resources and
5 Results and analysis
5.7 Analysis of identified challenges, resources and capabilities

to be developed capabilities on how to best overcome these challenges. An overview of the
identified challenges on all four levels along the value chain is given below in figure 6.

As can be seen from figure 6, at each step of the data value chain different challenges
occur. While in the beginning during data discovery eleven challenges have been identified,
data integration poses only four challenges to overcome. Finally, during data exploitation six
challenges must be overcome. Concluding, the findings show that more than 50 per cent of
all challenges are currently seen during data discovery stage and challenges do not occur on
every step in the value chain on all four levels. It should be noted that there was no
indication by the interviewees which challenges are the largest to overcome, therefore the
size of the boxes in figure 6 do not relate the magnitude of the challenge.

5.7 Analysis of identified challenges, resources and capabilities

The interviews yielded many challenges related to the discovery, integration and exploitation
of data in manufacturing companies. In total, 21 challenges were identified by interviewees
but, as the findings highlight, most interviewees had also possible solutions in mind of how to
overcome these challenges. These possible solutions are based on the acquisition of new
resources by the manufacturing company or the development of new capabilities by the
individual, managers or the whole organization. Therefore, this analysis of the results is
structured as follows: Firstly, the identified challenges will be evaluated and compared to
theory and concepts related to Big Data as they are described in chapter 2. Hereby, mostly
technical challenges will be covered due to the proximity of Big Data concepts to technical
aspects of Big Data. Secondly, the results will be analysed using the theoretical concepts of
the resource-based view in its two dimensions, static and dynamic as they are presented in
chapter 3.

Data value chain

In general, it can be derived from figure 6 that manufacturing companies follow the data
value chain as proposed by Miller & Mork (2013), since the interviewees identified
challenges along the three main process steps in the framework. These twenty-one
challenges will be analysed per each process step. Miller & Mork (2013) identified for the
steps of data discovery and data exploitation several sub-process steps. While the findings
5 Results and analysis
5.7 Analysis of identified challenges, resources and capabilities

indicate that during data discovery the sub-processes of collect and annotate and prepare are recognized, the interviewees were not able to point out the three proposed sub-processes in data exploitation. This may indicate that the interviewed manufacturing companies, technology suppliers and experts are currently more concerned with the data discovery and data integration before they think about the final step of data exploitation. This finding may indicate that manufacturing companies have already established a better view of the data discovery stage and are currently exploring this stage. Subsequently, the latter two stages are less mature and therefore have not that depth of detail as the first stage. Possible reasons to explain these findings are presented in chapter 6.2.

Data discovery
The data discovery stage is concerned with data volume, variety and value as stated in the previous analysis chapter. Interviewees outlined the challenge of the limitations of current IT infrastructure, which hinders the introduction of Big Data. This is in line with relevant and widely-accepted definitions of Big Data (Cukier & Mayer-Schoenberger 2013). For example, Manyika et al. (2011) stated that Big Data consist of “datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyse” (Manyika et al. 2011, p. 1). Moreover, the interviewees distinguished between internal and external data when getting access to data sources and collecting data. This supports the statement that for valuable and meaningful insights data from multiple sources must be analysed and further that data are also lying outside of the company’s physical boundaries (Davenport & Dyché 2013; Accenture Analytics 2014; Abbasi et al. 2016).

Gaining access to data sources poses challenges on two levels, namely at first on the technical level on how to technically connect to the data source and at second on the managerial level on how to negotiate and ensure access on a governance or political level. This is in accordance with the research by Klein & Verhulst (2017) who also stated these two challenges. Regarding the technical connectivity, the interviewees pointed out two challenges. First, the acquisition of additional internal data using sensors to integrate IT and OT, which is consistent with the reviewed literature, where the emergence of new technologies such as IoT and sensors can be retrofitted to production assets to establish connectivity to IT systems (Koch et al. 2014; Chong & Shi 2015). Second, enabling access to internal data from other departments as well as external data by integrating and networking all relevant IT systems. Hereby, integration of IT systems must occur on both levels, horizontal and vertical integration, which is in line with studies by Fraunhofer IPT (2015) and Schröder (2016). In the data discovery phase, interviewees also outlined that although it would be beneficial to acquire as much data as possible, too much data collection does not necessarily lead to more insights. Instead, a data strategy should be set up to determine which data should be collected for which business problems. This calls for the appreciation of data value and confirms the findings from previous research by Lavalle et al. (2010) and Fraunhofer IPT (2015).

The next step in the data value chain framework is the integration of data. The interviewees highlighted the organizational challenge of preparing and organizing the data prior integration, which is in line with the statement by Fisher et al. (2012, p. 55), who stated that the integrated data sets must be “organized, partitioned, and prepared before they are uploaded”. Since the integration of data also poses a challenge to the current IT systems, as was highlighted by interviewees before, one cited managerial challenges by interviewees was the necessity to ensure sufficient financial investments. This was also confirmed by Dijcks (2013) who argued that investments to upgrade current IT infrastructure must be made because otherwise IT systems will not be able to handle Big Data.

Data integration
During data integration, two technical challenges were identified, which are related to establishing a data structure and the creation of a powerful, flexible and scalable Big Data
5 Results and analysis
5.7 Analysis of identified challenges, resources and capabilities

platform. The importance of a data structure is also highlighted by Zhao et al. (2014) and Fisher et al. (2012). Especially Fisher et al. (2012) recommended that a data structure should be logical to provide clear data labels and efficient to reduce computation power for data processes.

The other challenge was to create a new Big Data platform. This is in line with the view by Chen et al. (2014) and Villars et al. (2011) who stated the need for new IT systems and technologies to manage Big Data. Especially the integration of many different data sources is poses a challenge for IT systems due to data quality concerns and therefore a Big Data platform should be able to maintain data quality (Kwon et al. 2014). The more diverse data sources shall be integrated, the more challenging it gets for IT systems (Abbasi et al. 2016; Gupta & George 2016). Since the interviewees talked about the integration of both, internal and external data, the challenges becomes even larger (Schröder 2016). At last, the interviewees emphasized the need for one central data platform, where all data are stored together in one place. This can be mapped to research by Xiang et al. (2016), which proposed the strategy of physical centralization of all data sources into one single data warehouse to act as one central data source for further data analysis.

Data exploitation

Overall, the findings from the interviews indicate that interviewees were aware of the importance of data analysis, however on a technical level they were not able to outline more detailed challenges than data analysis and visualization. In general, interviewees stressed the importance of the Big Data platform to not only be able to integrate data from multiple sources but also to process these data by analysing and visualizing them. This confirms the statements made by (Yin & Kaynak 2015; Sivarajah et al. 2017) that even BDA is only a minor step in data exploitation, but it still requires new IT systems and new approaches for data analysis. Moreover, the interviewees emphasised the change in working processes and working culture when it comes to data analysis and visualization. These findings can be backed with research by (Russom 2011; Chen et al. 2012; Kwon et al. 2014) who found that BDA significantly differs from traditional BI in terms of statistical analysis approaches and variety of data sources.

The aspect of data visualization mentioned by the interviewees is another important step, since only then data insights can be presented in such a way that employees can easily understand the information which was revealed by data analysis and act accordingly (Davenport & Dyché 2013). However, applying this new information of data visualization to improve business operations was merely stated by interviewees. Only one interviewed manufacturing company (MC1) used BDA results as a basis for data-driven decision making, however in the beginning employees did not want to trust the analysis results. On a positive note, the interviewees understood the need to analyse data for value creation, since only collecting and storing data is valueless (Sivarajah et al. 2017). However, as Russom (2011) pointed out, it is difficult to apply insights gained from data analysis to business operations. This may also explain why interviewees were for now not concerned with data-driven decision making (Tien 2013; Abbasi et al. 2016). One possible explanation for this is given by Sivarajah et al. (2017) who stated that retrieving the value from data analysis is not possible without human intervention and therefore often the management and business skills to apply Big Data insights lags behind. Thus, as proposed by Provost & Fawcett (2013) the interviewed manufacturing companies did not follow for now effective data science, since it would require the application of a data perspective to underlying business problems.

What can also be analysed from the findings is the degree to which the interviewed manufacturing companies were already exploiting data according to the stages of the progress of application and usage of BDA by Vidgen (2014). While Vidgen (2014) proposed six evolutionary stages in total, the research findings indicate that currently only the first two stages of Vidgen’s stage model are applicable to the research findings. The findings from
5 Results and analysis
5.7 Analysis of identified challenges, resources and capabilities

Interviews confirmed that for now only fragmented or localized usage of BDA is performed. For example, the German electronics manufacturer performs BDA at one production line only and is therefore according to Vidgen (2014) at the first stage. The adhesive manufacturer (MC1) already sourced data from its suppliers and based some of their decisions on the production floor on the results of data analysis. Thus, it can be concluded that this company already was in the second stage of localized BDA applications and further hat some elements from the fourth stage, since they made some data-driven decisions. However, the third stage of Vidgen’s stage model was not touched by this company, since there was no indication of the establishment of a central analytics department. Interestingly, MC4 was in very early stages when it comes to BDA, but this company already created a dedicated department and new employee roles as coordinator and negotiator between the IT department and production department. Therefore, this company shared characteristics with the third stage of Vidgen’s model.

Resources and capabilities
In order to overcome the challenges, the interviewees often talked about possible capabilities and resources they acquired or developed. Resources and capabilities, which were suggested by the interviewees, can be divided tangible resources (e.g. IT systems, Big Data platforms, internal and external data and financial investments), intangible resources (e.g. organizational and cultural change and change, cross-functional cooperation and coordination) and human resources (e.g. new skills, new talent, managerial support and leadership) as proposed by Grant (1991).

Since interviewees stated that a Big Data platform depends on the limitations and architecture of current IT systems, it is very likely that each company will have a unique Big Data platform tailored to their specific needs according to the Big Data strategy. Therefore, this kind of resource can be seen as difficult or even impossible to imitate and therefore the imitability aspect of a resource holds true (Teece et al. 1997). The challenge to find an implementation partner leads, if successful, to a long-term partner, who helps the company implementing Big Data initiatives by providing expertise and knowledge from outside. This long-term partner can also considered as a resource according to the definition of Helfat & Peteraf (2003), since access to this resource would occur on a semi-permanent basis. Nevertheless, for this resource the imitability aspect would not hold, since other firms can also contract the same firm as a partner for implementation. Furthermore, interviewees viewed a wide variety of internal and external data sources as a resource too. This may hold true for internal data which the firm generates during their company-specific business operations and external data sources which are sourced from business partners such as suppliers or customers. However, data coming from data providers or even freely available public data can be acquired from every company. Therefore, according to Teece et al. (1997) some external data cannot be seen as resources since they are not firm-specific assets. It can be also argued that data as a resource in general does not fulfil the "rare" criteria, since the volume of Big Data is large. Nevertheless, when considering all data in the central data warehouse at the Big Data platform after both internal and external data have been integrated as a resource, it can be said that each company will most likely have a company-specific data repository, unique in data volume and data variety and each data repository is valuable only for its owning firm.

When it comes to capabilities, research suggested that each capability consists of one or more different resources (Grant 1991). Further, research characterized capabilities as multidimensional and structured as well as they require coordination efforts between the company’s organization and other resources (Winter 2003; Grant 1991). These characteristics of capabilities can be also found in this research findings, since suggested capabilities by the interviewees require multiple resources from different levels. For example, the capability of collecting data from external data sources requires resources on the technical level (connectivity and networking of IT systems) but also managerial resources to
5 Results and analysis
5.7 Analysis of identified challenges, resources and capabilities

Negotiate access to these data sources. Another capability cited by the interviewees was the ability to perform organizational and cultural change. This capability also touched upon three levels, the individual, organizational and managerial respectively. While on the organizational level the organizational data-driven culture is classified as an intangible resource, on the individual and managerial level human resources are needed. While managers need to raise awareness for the need of organizational change and later must provide managerial support and leadership to facilitate this change, on the individual level employees need to develop new skills related to the appreciation of data value, to overcome data security and privacy issues and finally to trust data insights. Only if all these resources and capabilities on all three levels are coordinated together, the capability of an organization to transform its culture towards a data-driven culture can be achieved. All these examples provided above are capabilities, since according to Helfat et al. (2007, p. 1), a capability can describe a task or activity performed by an organization or firm.

In order to achieve competitive advantage with Big Data, managers of manufacturing companies have two main tasks according to Makadok (2001). These two are resource picking and capability building. When applying this to the research findings, it becomes clear that in the beginning, when managers define a data strategy, they also decide on which resources need to be picked and which capabilities need to be built where. For example, decisions need to be made which Big Data platform suits the company best and which implementation partner (as an external resource) is chosen. Moreover, the data strategy further decides on which data are needed and therefore it needs to be decided which capabilities are needed to identify valuable data sources, gain access to these data sources, collect these data, integrate data and analyse them. This is also similar to defining a business strategy in the RBV, since the data strategy also determines where and which resources and capabilities related to Big Data need to be acquired or built up in the long term (Teece et al. 1997).

When mapping the research findings to the capability lifecycle by Helfat & Peteraf (2003), the following can be analysed: Interviewees stated that in the Big Data era employees need to develop new skills. Some interviewees raised concerns that the employee’s new skill set is focused on Big Data, IT systems and digitalization and will replace mechanical engineering specific knowledge. According to the capability lifecycle it would mean that the mechanical engineering capabilities by employees which are currently in the maturity stage will either retire or will gradually decline over time (retrenchment). However, during the data discovery stage those mechanical engineering competencies are needed to determine which additional data must be collected to gain more insights. Interviewees argued that it requires deep technical expertise and extensive process knowledge to apply a digital data perspective to physical production processes. In other words, these mechanical engineering competencies can be also replicated or redeployed for determining which additional data are there to capture. Moreover, interviewees stated the need for a capability of managers to negotiate data access with business partners, for example for customer or supplier provided data. Some interviewees said it is beneficial to have strong and long-term committed relationships with business partners as it will facilitate easier data access and data sharing. Therefore, it can be argued that capabilities related to managing and maintaining business partner relationships will be renewed, since exchanging data for BDA can strengthening these partnerships and may lead to even longer lasting relationships.

Dynamic capabilities
Interviewees suggested that there is a chronological sequence of how a company implements its Big Data initiatives. The research findings indicated that manufacturing companies followed the data value chain by Miller & Mork (2013), which consist of three sequential steps. Secondly, interviewees stated that manufacturing companies are at first applying an internal view when it comes to the discovery, integration and exploitation of data and as a second step collecting and integrating data from external sources. As a result, DC
5 Results and analysis
5.7 Analysis of identified challenges, resources and capabilities

occur during the implementation of Big Data and are therefore applicable here (Ambriosini et al. 2009). As Teece (2007) outlined, changing technological opportunities, such as Big Data, require to adapt a firm’s resources and capabilities. A firm’s resource base needs to be developed, altered or changed to fill gaps in its current resource base when emerging technology trends lead to a change in business strategy (Grant 1991). Thus, once managers have understood the need for change towards Big Data and developed a data strategy, the process of altering the firm’s current resource base begins to acquire new resources and develop new capabilities to address and close current resource gaps. For example, once the data strategy defines the integration of internal and external data sources, the current IT systems are not sufficient resources anymore and need to be replaced by a new, more capable Big Data platform. Hence, once a data strategy is defined and the management outlined a clear course of action, the firm develops DC as it is extending and modifying its resource base on purpose (Helfat et al. 2007). Hereby, managers must look at modifying internal resources and capabilities, which the firm already owns or controls but also to think about acquiring resources externally (Grant 1991; Teece et al. 1997). The research findings indicate at least three resources which can be acquired externally. First, external data sources from either business partners, data providers or public data, second, knowledge and expertise from a technology consultant firm as an implementation partner and third, new skills by hiring new employees who are experts in Big Data.

The DC found in the research findings can also be categorized along processes, positions and paths as proposed by Teece et al. (1997).

Processes
As proposed by Teece (2007) DC can be divided into three capacities, which were to sense and shape opportunities, seize them and maintain competitiveness through strategic change of the resource base. The findings indicated that in the stage of data discovery the managerial challenges related to understand the need for change and define a data strategy while organizational challenges related to become open for change and lastly individual challenge was to appreciate data value. During this stage, the DC to sense and shape opportunities can be perceived. All identified challenges refer to coping with a newly emerging business opportunity, namely to acquire Big Data for enhanced decision making and business improvements. Thus, as a first step, it is the task of the management to apply sensing and opportunity shaping to prepare for the company for Big Data. Seizing these opportunities becomes the next logical step, as was also indicated by the findings from the interviews. Challenges here related to understand which data sources are needed, how to get access to them and finally to integrate them. Hereby, internal resources such as the IT systems need to be developed further towards a capable Big Data platform but also capabilities such as negotiating with supplies for data access are enhanced and strengthened. During the last step of the data value chain model, data exploitation, the interviewees indicated that there will be changes to the way of working and new skills of employees will be necessary to perform data analysis, interpret the results and apply the gained insights to business operations. This might apply to the last capacity of DC, as companies strategically change and reconfigure their resources and capabilities to build a future resource base. According to Teece et al. (1997) and Bowman & Ambrosini (2003) DC can be further divided into the four sub-processes. When mapping these four processes to the research findings, the processes of reconfiguration, learning and integration can be found. For instance, the current IT systems as physical resources need to be transformed to a powerful Big Data platform and for effective implementation and usage need to be combined with resources from the outside, namely the implementation partner. Here, the process of learning can be also observed, since according to Ambriosini et al. (2009) includes experience gain and the ability to learn and execute things more efficient. Thus, the DC to learn from the implementation partner is important here. Of course, when the company is hiring new Big Data talent from outside, this process applies also here. At last, the findings indicate that managers need to be able to form a new resource base of the Big
5 Results and analysis
5.7 Analysis of identified challenges, resources and capabilities

Data era by coordinating existing resources and capabilities such as employees, organizational culture and processes as well as internal data and current IT systems to form the resource base of the Big Data era, which consists of new organizational roles and structures, a data-driven culture, Big Data platform and highly skilled employees.

Positions
As the interviews with the five different manufacturing firms indicated, each company takes a slightly different approach when it comes to implement Big Data into the organization. For example, depending on which business problems there must be solved with the help of Big Data, companies take different approaches in upgrading IT systems and deciding which additional data need to be captured and analysed. Which data can, are and will be collected depends here first on the negotiation power of the company and second on their relationships with suppliers and customers. Interviewees indicated that negotiation power increases if there are mutual benefits promised and strong and long-lasting relationships with business partner increased the likelihood of receiving data access. Further, the interviewed manufacturing companies applied varying approaches in changing the organizational layout to reflect their Big Data needs. While some companies created new roles and organizational departments, others relied or will rely on the help of external partners entirely. The magnitude of changes is of course somehow limited in the ability of freeing up financial resources.

Paths
The research findings imply that there are path dependencies as they are defined by (Teece et al. 1997, p. 515). The concept of path dependency can be related to two challenges, defining a data strategy and deciding upon a data structure for the Big Data platform. First, the challenge of developing a data strategy decides for which purposes which data need to be captured, collected and analysed. Subsequently, depending on the data types and sources specified in the data strategy, a Big Data platform technology and its data structure is selected. However, building a data structure and define rules on how to organize the data needs to be determined upfront and cannot be changed easily later. Therefore, the decisions being made during the building of the data structure will limit the company's future choices in integrating additional data types or data from sources, whose data structures do not fit the data structure of the company's Big Data platform. Moreover, insufficient volume of financial investments being allocated to Big Data initiatives might also limit future choices to some extent. At last, the company's current IT systems and organizational layout might also influence the number of future paths open to the firm, as both pose limitations to Big Data initiatives. For example, not having the willingness to hire new talent from outside or allocate sufficient financial funds to upgrade IT systems on large scale will have a limiting impact on any Big Data initiatives.

At last, Wang & Ahmed (2007) proposed a different set of DC, adaptive, absorptive and innovative capability respectively. According to these three capabilities, the research findings reveal the following. Adaptive capabilities can be observed during the data discovery stage, since one challenge on the managerial level related to understand the need for change and subsequently prepare the organization to change. Since adaptive capabilities describe the capacity to align a firm's resource base to cope with changes in the environment, the challenge of defining a data strategy can also be linked to adaptive capability. Thus, the decision to adopt the new technological concept of Big Data as a new lever of operational improvements has a significant impact of changing internal organizational resources (Wang & Ahmed 2007). Absorptive capabilities can also be derived from the research results. For instance, the interviewees named the challenges of integrating external data to combine them with internal data for data analysis and to find a partner for implementation as well as to hire new talent from outside the organization or to acquire new skills. These challenges are all related to the acquisition and usage of outside knowledge for Big Data initiatives. For example, the implementation partner brings expertise in building and maintaining Big Data
5 Results and analysis
5.7 Analysis of identified challenges, resources and capabilities

platforms and newly hired talent has knowledge about data analysis and data science. Hence, to use this new acquired knowledge, the organization must be able to learn and combine it with already available internal knowledge to build new expertise. In other words, absorbing new knowledge is important for successful implementation of Big Data. At last, innovative capabilities do not only refer to the development of new products or services, but also to change processes within an organization. The findings revealed that change is an important challenge of Big Data implementation, since not only technological change must be coped with but also change occurs on the organizational and individual level while the facilitation of change occurs on the managerial level. Furthermore, the interviewees did also mention the different purposes of data collection and the value of data. At the end, each Big Data project results in value creation, whether value is created on bottom line improvement or products are enhanced or totally new developed. Thus, during the last stage of data exploitation, companies must have innovative capabilities to apply the outcomes of data analysis for value creation.
6 Discussion
6.1 Research findings in relation to data sources

In this chapter, the results obtained from the interviews are discussed in relation to previous research findings from other authors in the field. Further, the results are discussed on a higher level and from a wider perspective. Finally, the aspect of sustainability as well as ethical considerations and risks are outlined.

6.1 Research findings in relation to data sources

Based on the findings of this research, four topics regarding data sources for Big Data are worth discussing:

1. Data collection strategy

According to the interviewees, strategies for data collection follow a value-driven approach rather than a volume-driven approach. Hereby, the use case defines what the company wants to achieve with the data, whether it is to optimize the process, to detect failures better, to monitor quality or to increase the overall production process. This can heavily vary from use case to use case. Depending on the desired aim and outcome of the use case, data are collected. It is important that for each use case one needs to identify the data needed and ask whether these data are already available. This finding is confirmed by Lavalle et al. (2011), as their research revealed that placing emphasis on collecting all kinds of data before determining the reason or potential benefit does lead to suboptimal outcome as well as wasted resources. Instead, companies should think firstly about use cases and questions, which shall be answered with Big Data and then collecting the specific data needed for these purposes. As mentioned before, both internal and external data are considered for BDA to achieve a holistic data perspective. This finding is consistent with the results of earlier research investigating IT system integration of digital supply chains (Rai et al. 2006). The study emphasized the need for internal and external data and IT system integration to overcome the limitations of fragmented IT systems and data silos of manufacturing companies. However, Zhao et al. (2014, p. 173) pointed out that “selecting the relevant internal and external data features will require human expertise”. Thus, an efficient execution of the data collection strategy will take time until individuals gained first experience in understanding the value of data.

2. Sequence of utilizing data sources

As the results indicate, a first move towards data utilization must start with those data, which are already in place. Those will mostly be internal, structured data from the main IT systems. These data should firstly be used to gain insights and later more data sources from both new internal sources as well as external data can be added. This is to some extent consistent with previous research findings. On the other hand, Lavalle et al. (2011) suggested that at first companies should use already available data to conduct first data analyses. Results from these analyses shall be used then to identify data gaps indicating where and which data need to be collected. On the other hand, Kwon et al. (2014) concluded that the greater the positive user experience while working with internal data sources, the less likely is the collection of external data for BDA (Kwon et al. 2014). A possible explanation is that companies rather focus on a certain skill sets and do not want to expand on it (Kwon et al. 2014). In contrast, when a company had already accumulated significant experience with external data, it is more likely that the company intends to explore internal data for analytical reasons. Thus, positive experience with external data sources can facilitate the intention to acquire internal data sources, too (Kwon et al. 2014). Despite these findings being interesting, they are not reflecting the sequence of firstly exploring internal data and secondly adding external data. Manufacturing companies would not desire to acquire external data sources, when they made positive experiences with the exploration and exploitation of internal data. Likewise, when a manufacturing company encounters positive
experience with the usage of external data, it is more inclined to focus on internal data, too. This misalignment might be due to the following reasons: The surveyed companies only reflect the same industries to about 20 per cent. Moreover, the geographical scope of their research was South Korea, which probably introduces cultural differences of the interviewed companies. In this research the interviewees stated that manufacturing companies already operate in very data-rich environments. To firstly explore and exploit already owned and self-generated data is thus very likely compared to acquiring data from across the company boundaries. Furthermore, experimenting with internal data requires less financial investments and only maneuvering within intra-company politics in contrast dealing with more complex security barriers and data ownership questions when it comes to external data. To conclude, for manufacturing firms that have no or only made little progress with BDA so far, the focus of internal data and vertical data integration presents a more fruitful starting point for quick wins and small success stories to fuel their appetite to progress further with more data sources.

3. Access to and control over data sources
Data sources are either under full control of the company in vertical integration or horizontal integrated with less control to increase transparency. In relation to the control of data and data sources, there is obviously a need to coordinate and cooperate in order to get data required from all different areas. It is not easy to just collect data; it is difficult to convince data owners to get the data and tell them why the data are needed. It is beneficial to have good relationships with suppliers, which enables data access and exchange. Therefore, it is crucial to have a good network both internal and external to the company to ensure good communication to get data access and data exchange, to operate across departments and emphasize on the importance of mutual benefits for both sharing parties (Klein & Verhulst 2017).

4. Data as competitive advantage
Open data cannot be of a competitive advantage, since these data are available to everyone and are therefore not fulfilling all VRIN criteria proposed by Barney (1991). This issue was also outlined by Braganza et al. (2017), who claimed that data as a resource for competitive advantage is not possible. However, manufacturing companies can create competitive advantage if they merge open data with their internal data. Competitiveness based on data cannot rely on external data, which are accessible for everyone. That is why internal data need to be combined with external data to have another basis for decision making which is different from competitors.

6.2 Research findings in relation to Big Data challenges
When comparing the challenges on the technical, managerial, organizational and individual level with findings from previous studies (chapter 3.3), most of the challenges which were identified by the interviewees can be also found in the literature.

Technical challenges
Previous research also highlighted the limitations of current IT infrastructure, the challenge to process data and establishing access to data. Interestingly, other technical challenges mentioned in the literature such as data policies, data quality, data security and integration of data were covered by the interviewees in two main challenges: The first one is to develop and maintain a certain data structure to integrate data in an efficient way. This data structure must be well planned and encompass a holistic data perspective, since Kambatla et al. (2014) highlighted that an inclusive analytics framework needs to be able to integrate data in one organized structure from all internal and external sources, for example supply chain management, customers, sales, marketing etc. The second challenge was to build or create a capable Big Data platform, which should be flexible and scalable. This is consistent with
6 Discussion
6.2 Research findings in relation to Big Data challenges

the findings from Wamba et al. (2017), who determined that a firm’s BDA capabilities is influenced by infrastructure capability which consists of the three pillars of infrastructure connectivity, compatibility and modularity. The aspects of connectivity and compatibility are coherent with the other cited challenges of technical connectivity to data sources and the lack of standards to connect different machines and IT systems. Regarding the aspect of modularity, interviewees indicated that most likely current IT systems will be sequentially upgraded or replaced in a phased IT transformation and therefore old and new IT systems must harmonize. In addition, modularity promotes and supports the scalability of the Big Data platform. However, Chen et al. (2013) concluded that building a perfect Big Data platform from scratch is unmanageable and nearly impossible to achieve. Due to the fast-growing dimensions of volume, velocity and variety of Big Data, the authors propose to build a limited functioning platform with minimal service and from there incrementally upgrade the Big Data platform, using time, money and expertise, to its full functionality.

While previous research categorized data quality, data policies and data security as technical challenges, the findings from the interviewees indicated that these challenges are more related to managerial challenges and individual challenges. One explanation could be that technology advances faster than humans can adapt (Barlow 2013) and therefore challenges such as data quality and data security are nowadays not technical issues anymore, but these challenges need to be overcome by changing the individual. Novel are the more detailed challenges of gaining access to both internal and external data sources, which differ from each other. The findings revealed that sensors and new connectivity solutions could integrate machines and IT systems vertically within a company. To connect with business partners such as suppliers and customers, IT integration within the horizontal value chain is necessary. This is in line with findings by Fraunhofer IPT (2015) and consultancy reports by Koch et al. (2014) and Rinn & Blanchet (2014) as well as Olama et al. (2014). Besides the establishment of data accessibility, Olama et al. (2014) proposed a common data platform and most importantly a consolidated data model as important features for successful data integration. Hereby, the description of the consolidated data model is consistent with this research’s findings; since it aims to provide a logical data structure to universally understand all data elements and a common data label dictionary. At last, interviewees cited the technical challenge of data analysis and visualization. Research by Philip Chen & Zhang (2014) concluded that most companies have to deal with these kinds of challenges such as data capture, searching, analysis and visualization. Bitkom e.V. et al. (2016) developed implementation guidelines for the use of data and data analysis in manufacturing companies, which addresses the challenges of data access, data integration, data structure and data analysis too.

Managerial challenges
Findings from this study and the scholarly literature overlap as both highlight managers to define a data strategy (including data governance), provide financial investments and provide managerial support and leadership to enable cultural and organizational change. Indeed, previous research concluded that large investments are required to finance new IT hardware and software equipment (Sivarajah et al. 2017; Intel Corporation 2014). Still, these substantial investment have to be combined with the uncertainty about the benefits are big challenges for companies (Lorenz et al. 2016). In this regard, Gupta & George (2016) recommended that although expected results and return of investment will not be seen immediately, it is advised “to be persistent and devote enough time to BDA” (Gupta & George 2016, p. 1052) until first results are achieved. In addition to financial funds knowledge has to be acquired, for example in model factories, where practical use cases can be quantified in both implementation and benefits (Lorenz et al. 2016). If companies succeed, they can achieve up to three per cent faster productivity when having invested in new IT systems (Tambe 2014). Thus, managers should consider the value of new Big Data technologies and their associated costs (Tambe 2014). Managerial support and leadership is also important, since companies face uncertainty and risks when new technologies require a
6 Discussion
6.2 Research findings in relation to Big Data challenges

paradigm shift (McKinsey Digital 2015). Indeed, earlier research has concluded that absence and lack of support from management is one of the major issues why Big Data implementation in organization fail (Lavalle et al. 2011). Thus, strong top management support is essential in creating a new corporate culture open to innovative approaches with Big Data (Kiron et al. 2013).

Yet, not only managerial support is necessary but also managers’ leadership, skills and capabilities must be aligned with Big Data. Wamba et al. (2017) defined Big Data management capability consisting of four processes, which relate to planning, financing, coordinating and controlling of all Big Data efforts. Gupta & George (2016) related new managerial skills to the ability to apply the results and insights gained from BDA to certain business situations, processes or needs. Managers “should have the ability to understand the current and predict the future needs of other business units, customers and other partners” (Gupta & George 2016, p. 1053). This reflects the managerial challenge of creating a data strategy, since a data strategy not only defines which data shall be collected from which sources, but also which overall business goal shall be achieved with BDA. This is in accordance with Lavalle et al. (2010) who concluded that to start an analytics project there has to be a clear business goal determined to analyse data to answer open questions, not vice versa. Likewise, the importance of strong working relationships and achieving trust to benefit organization-wide from Big Data efforts was studied by Gupta & George (2016). This is related to the challenge of preparing the organization for change and facilitating change. Interviewees outlined that it is the manager’s task to guide cultural change.

This study discovered two challenges managers face, which are firstly the role they play in negotiating data access to both internal and external data sources and secondly their task to find a partner to support the company in their Big Data initiatives. Klein & Verhulst (2017) claimed that the likelihood of sharing data increases if there is a clear incentive for each sharing party or mutual benefit. Hereby, the clarity of incentives and sharing of risks will facilitate access and enable data integration from external sources and business partners. Scientific studies about implementation partners are scarce. Vessey & Brown (2014) stated in their research that for large, complex IT implementations, third-party consultants are often used as implementation partners to provide their expertise in managing such projects, despite paying high fees for these services. However, their expertise is invaluable given high degrees of customization and process innovation to best serve the client’s needs.

Individual challenges
The challenge to acquire new skills was discussed in previous literature. For example, Davenport & Dyché (2013) raised concerns that lack of skills to perform advanced analytics hinder business decisions with BDA. Similarly, Chen et al. (2013) stated that there is need to acquire new skills related to the adoption of Big Data. Gupta & George (2016) investigated the difference between traditional IT capabilities and Big Data capabilities: While IT capabilities are concerned with supporting and enabling the daily operations of the firm, Big Data capabilities focus on the extraction of new insights from a large variety of data coming from many different sources. Thus, the responsibilities and skills of employees working with Big Data are fundamentally different from employees working in the IT department and therefore new skills need to be acquired. Wamba et al. (2017) concluded that newly learned skills related to Big Data capability of employees consist of technical knowledge, technological management knowledge, business knowledge and relational knowledge in order to understand Big Data, perform BDA and lastly apply the gained insights to solve business problems. These conclusions are similar to the findings by Chen et al. (2012), who viewed Big Data related skills as interdisciplinary and should comprise of IT skills, business skills but also soft aspects such as communication skills. Besides, Big Data skilled employees have to be competent in understanding the relevant business problems and develop suitable analytical methods to solve them. At last, Sicular (2012) claims that there is no certain, fixed skill set for Big Data, but rather Big Data skills have to be tailored to each
6 Discussion
6.2 Research findings in relation to Big Data challenges

respective Big Data project and should comprise multidisciplinary teams to have a broad range of skills.

Contrary, in this study three critical data-related challenges, which individuals have to overcome were identified. Those were firstly the appreciation of data value, secondly data security and privacy concerns and thirdly the creation of trust in insights revealed by data analysis. Regarding the appreciation of data value, Zhao et al. (2014, p. 173) outlined that "selecting the relevant internal and external data features will require human expertise". Hence, understanding the value in data and knowing which data to select for a certain use case is built up as expertise over time. While data privacy and security aspects have been widely discussed in literature, for example Chen et al. (2016) and Zhou et al. (2014) were among the few authors who discussed the issue of creating trust in data. They claim that for data to be trustworthy, the data source has to be trustworthy. Just adding more and more data from different sources is not necessarily increasing the value and trustworthiness of data. In addition, the authors emphasize the aspect of intuition importance. In their view, "computational intelligence should be tied with human intuition", meaning that even though data will reveal new insights, those results should be never fully trusted and thus human intuition and field knowledge will remain important to judge the analysis results.

Organizational challenges
The findings are consistent with previous research. This research findings and the literature review identified organizational and cultural change as challenges to be overcome. Adopting a data-analytical mind-set is a central aspect for the whole organization (Davenport 2006; Provost & Fawcett 2013). The corporate culture is affected by instead making decisions based on intuition now changing to data-driven (Abbasi et al. 2016). Hereby it is important to establish a data-driven culture not only at the top-management level but across all hierarchical levels in the organization since all employees should be able to base their decisions on well-grounded data evidence (McAfee & Brynjolfsson 2012; Gupta & George 2016). Several authors (e.g. McAfee & Brynjolfsson 2012; Lavalle et al. 2011; Ross et al. 2013) related the forming of a data-driven culture as an intangible resource. It was also viewed as a critical resource for the successful implementation of a Big Data project (Gupta & George 2016). McAfee & Brynjolfsson (2012) argued that the development of an organizational culture related to a data-driven mind-set is an important enabler to benefit from Big Data. Nevertheless, people have difficulties to change their decision behaviour from personal experience-based to data-driven (Lavalle et al. 2010). Lack of a data-driven culture is the main reason for the failure of reaching the objectives of Big Data projects (Lavalle et al. 2010). Maxwell et al. (2016) investigated the data-driven corporate culture in detail, which confirms many of this research’s findings. According to Maxwell et al. (2016), building a data-driven culture starts with the employee’s appreciation of data value and to have an organization-wide common understanding of the purpose of data-driven decision making. Achieving this common understanding has be the task of management executives and must occur before or while a data-driven organizational culture is created.

Further, changing work processes require new competencies and tasks, which can be either satisfied with education and training of employees or hiring talents from outside. In this matter, Philip Chen & Zhang (2014) stated that it will take time to develop and educate powerful Big Data scientists as they need both mathematical competences and business-related knowledge. Companies have the choice to either educate and train their own employees in those fields or hire new personnel (Lorenz et al. 2016). The challenge of cross-functional coordination and cooperation is covered in literature with the challenge of changing organizational processes (McKinsey Digital 2016). Research has also shown that companies benefit from a dedicated analysis department as a central point of contact for the whole organization when it comes to BI&A (Lavalle et al. 2011). Often organizational adoptions for the creation of a central data management department (McKinsey Digital 2016). As a data science department becomes central, a wide range of employees need to
interact with them and subsequently understand the basics of data-analytic thinking (Provost & Fawcett 2013). Not only managers need to base their decisions on data instead of intuition, but all employees should do this (McAfee & Brynjolfsson 2012). The wider the scope of using analytics for decision-making, the more likely companies are to perform well (Lavalle et al. 2011). Nevertheless, the organizational challenge of preparation and organization of data is quite new. Although Zhao et al. (2014) concluded that the selection of relevant data labels is of high importance, the process of how these data labels should be defined and managed was not further studied. In fact, to the author’s best knowledge, there are no published scientific articles about this challenge. Yet, the preparation and organization of data as an organizational challenge can introduce bias to the data and subsequently cause problems in data analysis according to Crawford (2013). She argues that as soon as a human factor is introduced to data (i.e. the organizational effort to create a data structure, data labels etc.), the data are losing their objectivity: “We give numbers their voice, draw inferences from them, and define their meaning through our interpretations” (Crawford 2013, para 2). Thus, human intervention during data collection and preparation will introduce biases to the data and will subsequently affect data analysis results. This further implies that the structure and organization of data also affects the employee’s trust in data.

Concluding, while many challenges identified within this research also have been outlined in previous research, yet some discovered challenges are new. By mapping the challenges along the data value chain, this study provides a novel process perspective and highlights challenges on each process step. The identified challenges, possible resources and capabilities to overcome these challenges touch upon the whole organization: Challenges occur on all hierarchical levels of the organization (managers as well as individuals as human assets), but also on organizational processes and touch upon the intangible aspects (organizational challenges such as culture). On a higher level are technical challenges, which interact and sometimes even determine the challenges which happen on the other three levels. Challenges on each level do not occur or happen in isolation. The interviewees indicated that challenges occurring on one level are also impacting or even creating challenges on another. For example, the change of corporate culture towards data-driven decision making does include the organizational challenges to become open for change and eventually perform the cultural change. Yet, it includes also the managerial challenges of understanding the need for change and facilitating the change process as well as the individual challenges of the appreciation of data value, overcoming data security and privacy concerns and finally trusting the insights gained from data analysis. This research finding is confirmed by Tambe (2014), who stated that adopting or implementing successfully Big Data requires changes affecting the whole firm: “Installing these capabilities often requires organization-wide changes to complement data-driven practices” (Tambe 2014, p. 1468).

Interestingly, it can be observed that most challenges have been identified in data discovery. In the two later stages of the data value chain less challenges have been named. This could have happened for two possible reasons: First, data discovery is the most important and thus most challenging step and therefore the most challenges need to be overcome at this stage. Second, this observation indicates that manufacturing companies currently have not been progressed further in the data value chain and are still focusing on the data discovery stage. Recent industry reports indicated that so far only few manufacturing companies have made substantial progresses in implementing Big Data in daily operations. Numbers range from 15 per cent (van der Meulen 2016) over 20 per cent (Lorenz et al. 2016) up to 56 per cent (McKinsey Digital 2016) of the observed and studies manufacturing companies which have successfully implemented first Big Data projects. Nevertheless, future outlooks are promising as over 80 per cent of manufacturing companies studied by Geissbauer et al. (2016) expect themselves to base their decisions on data analysis results within the next five years. The second reason can be backed up with multiple studies from consultancy firms, which concluded that most manufacturing firms have not made substantial progress in Big
6 Discussion
6.3 Research findings in relation to resources and capabilities

Data and only few firms have reached mature levels. In addition, the fact that only a few initiatives exist towards the analysis of Big Data is not surprising, since Hartmann et al. (2016) discovered, in their study about companies which rely on data for their business model, that only few of the observed companies performed advanced analytics.

6.3 Research findings in relation to resources and capabilities

In the following, the research findings related to resources and capabilities from the RBV are discussed in relation to previous research.

6.3.1 Static view

The findings from interviewees indicated that the proposed resources and capabilities could be classified into tangible, intangible and human resources. This is in line with previous research since Gupta & George (2016) concluded that building BDA capability is only successful and effective if firms acquiring resources on all three levels, tangible, intangible and human respectively. Interestingly, although proposed tangible resources are related to overcome technological challenges, the majority of proposed resources are classified as intangible and human resources to overcome non-technical challenges. Lavalle et al. (2011) confirmed this research finding, as the study concluded that even greater than technological concerns are organizational and cultural concerns when transforming into a Big Data-driven company. Generally, the research findings showed resources and capabilities to be developed across all enterprise levels, from the individual, managerial and organizational level. For example, changing the organizational culture touches upon all these three levels while the capability to gain access to data sources requires capability development on the technical and managerial level. Thus, the research findings suggest that resource and capability building do not occur separately on each level but are connected among the levels. This finding is in accordance with Davenport & Dyché (2013, p. 2) who indicated that the implementation of Big Data should not be an isolated topic, “but must be integrated with everything else that’s going on in the company”. Similar, Chen et al. (2012) concluded that for Big Data implementation not only technical aspects need to be considered, but softer organizational aspects such as developing new skill sets, data-driven organizational culture and communication of change are vital for success. Comparable conclusions were also made by Verhoef et al. (2016) who stated that companies can build Big Data capability by targeting four key areas: 1. develop and maintain the skills of individuals, 2. invest in IT infrastructure, 3. create organizational processes to reflect data-driven decision making and 4. change the organizational culture. A recent study by Wamba et al. (2017) arrived at a slightly different result. Eventually, their empirical study revealed that management capabilities are less important when it comes to Big Data, but personnel experience and skills were seen as most important to build Big Data capability. However, it should be noted that their study did not consider the aspect of organizational culture and therefore management capability was not seen as important. In contrast, this study’s findings discovered managerial resources and capabilities related to changing the organizational culture and therefore managerial capability is more weighted here.

6.3.2 Dynamic view

Besides the static view, the interview findings indicated certain dynamics in developing and building resources and capabilities over time. As aforementioned, these capabilities are called DC as they describe the change of operational capabilities over time (Winter 2003). The interviewees described different challenges when it comes to collect and get access to data. There is a difference in resources and capabilities between internal and external data. The interviewees specified how data discovery might occur, as they proposed to firstly concentrate efforts on utilizing internal data and later focus on external data.
6 Discussion
6.3 Research findings in relation to resources and capabilities

Based on the interview findings, classification scales by Lavalle et al. (2010) and Vidgen (2014) as well as inspiration of the matrices by Zhao et al. (2014) and Kache & Seuring (2017), the following matrix was generated, see figure 7. This matrix is called the ‘Big Data archetype matrix’ and is explained below.

**The Big Data archetype matrix**

It describes how manufacturing companies transform into a data-driven organization and the sequence of data collection, divided into internal and external data. This 2x2 matrix consists of four fields, of which each one describes an archetype. Each of these four fields is positioned along a vertical and horizontal axis. The horizontal axis refers to external data and the vertical axis refers to internal data. Both internal and external data are divided into existing and new data. Hereby, existing refers to data, which is already available and/or collected and new data refers to data, which needs to be accessed first before it can be collected. Depending on which data a manufacturing firm is currently collecting, it can be described with one of the four archetypes. The arrows between each archetype are indicating that it is possible to move between them. Manufacturing companies can become different archetypes when they progress in their efforts to collect data from new internal or external data sources. In the following, each archetype is defined.

**Definition of archetypes**

Each archetype differs in the degree of utilizing internal and external data sources for data analysis.

The first archetype (lower left corner) is labelled as ‘Status quo’. Before a manufacturing company starts any initiative towards Big Data, it makes ad-hoc use of analytics from fragmented data collected within the production department and only uses traditional BI techniques. Thus, current data exploration and exploitation can, according to Vidgen (2014), be characterized as fragmented. Compared to today’s possibilities of BDA, this archetype is rudimentary. A great potential of data, which are already collected, are not processed or further used. Moreover, the company is either neglecting all external data sources or is using them only to a small extent.

The second archetype (upper left corner) is labelled as ‘Vertical data integration’. The manufacturing company starts to search for new internal data sources, especially in the
6 Discussion
6.3 Research findings in relation to resources and capabilities

production department to establish a functional-wide emphasis on improving operations and margins with the help of data analysis (Vidgen 2014). First use cases are discussed, implemented and a data strategy is defined. Besides, first change initiatives are launched such as creating a central analytics service as part of a new organizational structure and laying the foundation for a data-centric corporate culture.

The third archetype (lower right corner) is labelled as ‘Horizontal data integration’. The manufacturing company has gained first experience with the analysis of internal data sources and is eager to explore more data sources outside of company boundaries. It begins to integrate data from multiple locations (e.g. supply chain, environment) and starts to exploit medium sized data sets (Vidgen 2014). In the supply chain, data can be acquired upstream (suppliers) and downstream (logistics, distribution, retail) to the company (Addo-Tenkorang & Petri T Helo 2016). Hereby, being granted access to data sources of suppliers or customers and technically connect to these new sources are main obstacles to overcome.

The fourth and last archetype (upper right corner) labelled as ‘Data ecosystem’. The company makes decisions throughout the whole organization based on evidence gained from data insights coming from a large variety of both internal and external data sources in the digital ecosystem (Vidgen 2014). By now, the manufacturing company has gained significant experience in the exploration and exploitation of data and has transformed processes and organizational culture towards being data centric. At last, an developed smart analytical IT system also supports the transition towards data-driven decision making (Lee et al. 2016).

Concluding, for a manufacturing firm the first step is to collect, store and utilize all internal data and then add external data from the supply chain and environment. All data sources contribute to the added benefit of data analysis. The goal is to have not only full data transparency of internal operations but to also have end-to-end SC transparency. By now it becomes clear that transitioning towards Industry 4.0 and company-wide data-driven decision making is an evolutionary, rather a revolutionary process (Kagermann 2015). How a manufacturing company can transition from one archetype to another is outlined in the following.

Transition between archetypes

In order to transition into another archetype, a manufacturing company needs to develop DC by changing its resources to meet new challenges, exploit new opportunities and adapt new technologies (Danneels 2002; Teece 2007; Ambriosini et al. 2009). DC enable the manufacturing firm to align its resource base to meet the new challenges each archetype demands to be overcome.

Starting from the lower left corner of the matrix, the manufacturing company has the archetype of ‘Status quo’. From there, it can connect, collect and acquire new data sources vertically or horizontally. While vertical direction refers to new data sources within the company boundaries, i.e. referring solely to internal data, horizontal data exploration and exploitation refers to data sources along the supply chain of the company. The interviewees indicated that manufacturing firms first pursue vertical data integration before focusing their efforts on acquiring data from external sources. When a manufacturing company has acquired substantial experience with the analysis of internal and external data sources and has aligned its entire business strategy as well as organizational layout towards data, it finally operates as the ‘Digital ecosystem’ archetype. On their journey to the ‘Digital ecosystem’, manufacturing companies must transition from ‘Status quo’ to ‘Vertical data integration’ to ‘Horizontal data integration’ to finally arrive at this last archetype. Therefore, three transitions must occur as indicated by the arrow numbers in figure 7.
6 Discussion

6.3 Research findings in relation to resources and capabilities

Dynamic capabilities for transition 1
For the first transition, it is necessary for manufacturing companies to understand the potential of data and establish first ideas what and how potential benefits can be captured. Creating a data strategy and understanding the need for change are likely challenges to overcome. Thus, as according to Teece (2007, p.1319), manufacturing firms must develop the “capacity to (1) sense and shape opportunities and threats and (2) to seize opportunities”. Sensing and sizing are important DC to understand the potential of Big Data and how to deal with this topic. However, it goes beyond of being purely innovative, as it involves realigning internal resources and organizational processes as well as structures to address these new opportunities. Further, as according to Wang & Ahmed (2007), another DC is required for this transition, which is adaptive capability. Similar to sensing and seizing capabilities, adaptive capabilities support exploration and exploitation of emerging opportunities and leads to new organizational forms (Wang & Ahmed 2007). Yet, this capability also includes the management’s task to understand the need for change, encourage employees to be open for change and finally facilitate change (Gibson & Birkinshaw 2004). To summarize, for the first transition there is internal flexibility required to align strategically internal resources with emerging business opportunities.

Dynamic capabilities for transition 2
As very similar to the transition 1, DC for transition 2 include sensing, shaping and seizing of emerging opportunities, but this time these capabilities are not internal focused but externally. In addition, since now the focus lies on the acquisition and collection of data from external sources, absorptive capability as a DC becomes necessary. Absorptive capability supports the manufacturing company to acquire new knowledge based in external data sources, integrate it with existing internal knowledge coming from internal data to create new knowledge as well as the collaboration with external business partners in this matter (Cohen & Levinthal 1990; Wang & Ahmed 2007).

Dynamic capabilities for transition 3
When a manufacturing company wants to operate in the digital ecosystem, the following DC are required to make this transition. At first, according to Teece (2007), firms need the DC of monitor and steadily maintain their competitiveness by developing and reconfiguring its resource base in alignment to its business strategy. Additionally, innovative capability is required to manage the transition. Staying on top of competition in a digital ecosystem requires constantly being innovative in both products and internal processes. Thus, innovative capability supports the processes to constantly change and innovate to be on top of new trends and opportunities emerging in the digital ecosystem.

In sum, it becomes clear that each transition requires a different set of DC, which include adaptive, absorptive and innovative capabilities as well as sensing and seizing mechanisms. Manufacturing companies must change their DC to manage the next transition. Only focusing on the development of DC is not sufficient, as these DC must change as well. Hence, higher-order capabilities that are able to modify DC also have to be considered (Winter 2003). On the one hand these higher-order capabilities are renewing DC, which modify and refresh the resource base, but on the other hand regenerative DC to alter the DC between each transition (Ambriosini et al. 2009).

Comparison to previous research
How do the concepts of the ‘Big Data archetype matrix’ relate to previous research? What first comes to mind is that the path towards the digital ecosystem consists of certain steps. As outlined before, this fact is confirmed by Kagermann (2015), who stated that the transition towards Industry 4.0 is an evolutionary, rather a revolutionary process. Thus, multiple evolutionary steps might be possible. The sequence of archetypes on the path towards the digital ecosystem can be compared to Kwon et al. (2014), which has, to the author’s knowledge, been identified as the only study investigating the sequence of adoption of
6 Discussion
6.5 Reflections on sustainability

internal and external data. When applying the findings by Kwon et al. (2014) to the matrix, the archetype 2 ‘Vertical data integration’ and archetype 3 ‘Horizontal data integration’ in the matrix should change. Possible reasons for the mismatch have already been outlined in chapter 6.1. Moreover, the matrix ends at the ‘Digital ecosystem’ archetype and views this archetype as the desired outcome. According to Lavalle et al. (2011) the desired state of a manufacturing firm must be this archetype, since the wider the scope of using analytics for decision-making, the more likely companies are to perform well. Companies who embrace data analytics to drive business decisions yield higher returns and are more productive than competitors (Manyika et al. 2011). The archetypes can also be compared to Lavalle et al. (2011) and their three archetypes (aspirational, experience and transformed) of how companies embrace BDA. The aspirational organization represents the first and second archetype, as they just have begun to explore BDA and utilize mostly internal data to improve processes or decrease costs. In addition, the necessary resources and capabilities for successful BDA are not yet in place. The experienced organization represents the third archetype in the matrix, as gained experience with BDA allows the firm to explore more data and new analytical analyses for further optimization. Finally, the transformed organization represents the fourth archetype, as firms operating in the ‘digital ecosystem’ archetype have substantial experience with BDA and aligned the whole organizational layout, processes, products with a data-centric business strategy.

6.5 Reflections on sustainability

The issue of sustainability is often connected to the term ‘triple bottom line’ which was created by Elkington (1994). Besides financial sustainability measures, Elkington added also measures which reflect social and environmental performance measures related to sustainability (Slaper & Hall 2011). This research presents guidance for manufacturing firms to acquire BDA. Thus, the aspect of sustainability of this research’s findings will be analysed based on these three measures.

Economic sustainability
Economic sustainability refers in a wider sense to the flow of money (Slaper & Hall 2011). As outlined before, the introduction of BDA into business operations can lead to bottom line improvements and revenue growth, which results in higher margins and profits. Efficiency increases will substantially contribute to the sustainable competitiveness of manufacturing companies (Koch et al. 2014). Furthermore, BDA can secure the long-term competitiveness against Asian manufacturing companies that try to undercut margins due to lower labour costs and reduces therefore the degree of outsourcing or offshoring. Considering future scenarios, lessons learned from the implementation of Big Data and achieving competitive advantage by protecting margins and foster business growth can help to prepare an organisation to finance the successful implementation of emerging technologies such as Blockchain (Bahga & Madisetti 2016).

Social sustainability
Social sustainability measures social capital, well-being and quality of life for communities and individuals (Slaper & Hall 2011). Big Data does not only promote social sustainability; it also poses a risk to it. On the one hand, increased competitiveness and long-term sustainability of manufacturing firms can lead to better wages and employment rates and thus increase well-being and quality of life. Moreover, Big Data demands new skills which could lead to improvements in education. On the other hand, BDA and smart analytical systems might take over certain job roles or make them obsolete. Moreover, social responsibility should not be neglected. Valuable knowledge and experience will be lost in the near future when a lot of older, experienced employees will retire (Lee et al. 2016). Their experience will be incorporated in smart analytical systems and artificial intelligence might be superior to human intelligence in these regards. Increased data capture and storage of
6 Discussion
6.5 Reflections on sustainability

personal data from both employees and business partners leads to issues of data ownership and data privacy. The increased number of employee-related data being captured, monitored and evaluated increases not only the transparency of working habits but also the employee’s private life. In this matter, the protection of employees and ethical considerations must be discussed on governmental level to formulate policies and guidelines. The aspect of data privacy, security and ownership is of high importance when it comes to using cloud technologies to store data with especially high concerns in Germany (Martucci et al. 2012; Dillon & Vossen 2014). This is due the general view that clouds are perceived as less secure than internal IT infrastructure (Kaufmann 2010). Thus, education about cloud benefits, improvements in cloud security, data privacy and data ownership are actions to follow to promote social sustainability in regards to Big Data.

Environmental sustainability
Environmental sustainability describes the indulgent consumption of natural resources, environmental protection and waste reduction (Slaper & Hall 2011). Possible improvements of BDA in manufacturing are operational efficiency, process innovation, and environmental impact (O’Donovan et al. 2015). Big Data helps firms to identify optimal production levels and reduces forecast errors in the supply chain which reduces waste from overproduction. Moreover, energy consumption can be decreased when machines are running at optimal levels. These factors contribute to higher resource utilization and hence a reduction of the overall number of resources needed for the same production volume.

To sum up, the introduction of Big Data in manufacturing companies does have a positive impact on economic and environmental sustainability. However, Big Data might have a negative impact on well-being and quality of life of individuals due to increased data transparency of personal data. Thus, ethical aspects need to be considered to protect employees and promote social sustainability. The introduction of new technologies such as Blockchain might solve this problem and is therefore a possible way forward.
7 Conclusion and recommendations

7.1 Reviewing the purpose and answering research questions

In this chapter, the main research findings are reviewed and summarized. Additionally, research contributions are outlined on theoretical and empirical level and managerial recommendations are formulated. Lastly, research limitation mitigation strategies are discussed and avenues for future research are suggested.

7.1 Reviewing the purpose and answering research questions

The purpose of this thesis was to investigate which resources manufacturing companies have to acquire and which capabilities must be developed to overcome current challenges related to the implementation of Big Data analytics by utilizing internal and external data sources. The aim was operationalized by the following research questions, which were answered with semi-structured interviews in chapter 5 and discussed in chapter 6.

RQ1: Which (new) internal and external data sources and types are considered by manufacturing companies for Big Data analysis?

When it comes to Big Data, manufacturing companies consider both internal and external data sources. Internal data sources include already stored data in IT systems from different departments as well data captured by sensors and machines in the production department. External data comprise data provided by business partners along the value chain, data from environment and open data, which are either freely available or can be purchased. While internal data are mostly used to close current data gaps and improve internal processes, external data are utilized to enhance data analyses for more sophisticated data-driven decision-making.

RQ2: How can manufacturing companies successfully integrate data from internal and external data sources for Big Data Analysis?

c) Which challenges need to be overcome?

Manufacturing companies, which intend to launch Big Data initiatives encounter multiple challenges across the managerial, organizational, individual and technical level. Formulating a data strategy, initiating and facilitating organizational and cultural change as well as negotiating data source access are among the managerial challenges. While organizational challenges include the formation of new processes due to new organizational design, the creation of an organization-wide data structure and mastering cultural change to become a data-driven organization. Individual challenges contain understanding the value of data, overcoming data privacy and security concerns as well as building trust in the outcome of data analyses. At last, technical challenges refer to upgrading old IT systems to establish a powerful smart analytical system, which is capable of multi-level data integration from all kinds of data sources and perform accurate data analyses quickly.

d) Which resources must be acquired and which capabilities developed?

To overcome the abovementioned challenges, manufacturing companies have to acquire resources and develop certain capabilities to succeed in their Big Data efforts. Specifically, the following tangible, intangible and human resources are proposed:

- Large data repository consisting of internal and external data
- Big Data platform capable of integrating, analysing and visualizing large amounts of data
- Sufficient financial investments
- Skilled and educated employees as well as analytical talent
- Organizational layout shaped towards Big Data analytics
- Data-driven corporate culture
7 Conclusion and recommendations
7.2 Research contribution

These resources are the input to develop the following proposed capabilities:

- Management capability to define a data strategy, communicate and facilitate organizational change
- Technical capability to select an appropriate technology platform and implementation partner
- Technical capability to understand data value, identify relevant data and get access to data on both legal and technical level
- Organizational capability to define, organize and manage data structure
- Organizational capability to perform organizational and cultural change
- Organizational capability for cross-functional coordination and cooperation and changing work processes
- Individual capability to appreciate data value, overcome data security and privacy concerns and have trust in data insights

In addition, the following dynamic capabilities are necessary to conquer constant change and technical progress during the journey of implementing Big Data organization-wide:

- Capacity of sensing and seizing of Big Data's business opportunities
- Adaptive capability to understand the potential of Big Data and be able to respond to this opportunity
- Absorptive capability to accumulate external knowledge in form of data to enhance business performance
- Innovative capability to develop and market new products and services that might be created through BDA
- Regenerative capability to change and alter the set of dynamic capabilities to stay on top of constant change

Building upon the answers of research question 1 and 2, the main research question can be answered.

MRQ: How can manufacturing companies explore and exploit internal and external data for Big Data analysis to transform towards a data-driven enterprise?

Successful BDA supports manufacturing firms to achieve superior business performances and competitive advantages. However, at first several challenges have to be mastered by acquiring resources and develop capabilities on the managerial, organizational, individual and technical level. Manufacturing companies consider both internal and external data sources as inputs for BDA. The journey towards data-driven decision based on BDA can be structured into four archetypes. Firstly, manufacturing companies have to develop an understanding of Big Data as well as sense and seize its potential. Secondly, internal data sources have to be explored to gain experience with Big Data through organizational learning. Thirdly, external data sources have to be acquired from the environment or even business partners to enhance BDA results and build upon already made experiences. Lastly, the manufacturing company is part of a digital ecosystem, where all organizational processes, products and business strategy are centred on data. Hereby, manufacturing companies must additionally develop dynamic capabilities to manage the transition between archetypes.

7.2 Research contribution

To the author’s knowledge this is the first attempt to determine resources and capabilities in a context of data sources and the progress of Big Data initiatives of manufacturing companies over time. Implications of research findings can be drawn on two levels academic and managerial respectively.
7 Conclusion and recommendations

7.2 Research contribution

7.3.1 Theoretical contribution
This thesis advances the understanding of implementing Big Data in manufacturing companies and contributes to scholarly literature that examines Big Data from a resource-based view. Big Data is a growing research area and has mainly been covered from a technical research lens. Growth of Big Data research in management research only picked up recently and is still immature, hence leaving multiple research avenues to be explored. This study was conducted from an organizational viewpoint, investigating which resources and capabilities are required to succeed in BDA. The findings add new insights to answer “one of the most interesting questions in the field of big data research today” (Phillips-Wren et al. 2015, p. 465). Specifically, the following theoretical contributions have been made:

1. **Process perspective**: Research points out challenges, resources and capabilities along a process framework called data value chain. This provides a novel view, as it indicates when and where certain resources and capabilities are required and provides structured and detailed insights in contrast to other research.

2. **Expanding on Big Data resources and capabilities**: This research proposes new resources and capabilities and therefore adds to the still expanding research topic on this issue. Particularly, this thesis lists capabilities related to the acquisition and integration of internal and external data and thus provides a novel perspective.

3. **Data source archetypes**: Emphasis is placed on the location of data sources and different data types given by the classification into internal and external data sources. This led to the development of the Big Data archetype matrix, indicating which data types and sources manufacturing companies acquire over time and how.

4. **Dynamic perspective**: By adding the dimension of time, it was possible to explain how manufacturing companies move forward with their Big Data efforts over time. This perspective enabled the investigation of dynamic capabilities and therewith to obtain new insights of integrating different data sources over time.

7.3.2 Empirical and practical contribution
This thesis contributes to empirics by collecting qualitative interview data from representatives of manufacturing companies, technology suppliers and management consultants. The geographical area of data collection was limited to Europe, specifically the countries of Sweden and Germany. The collected data provide information about the status quo of Big Data implementation, encountered challenges and possible strategies to overcome them. Particularly, which data types and data sources manufacturing companies consider for their data analyses and how these are integrated and utilized over time. Thus, the empirical material can be viewed as a snapshot of the current view of Big Data in the European manufacturing industry.

This thesis also provides guidance for managers of manufacturing companies, which are about to start or have already started first Big Data initiatives. For each of the four archetypes mentioned above, the following managerial recommendations can be deducted from the research findings:

...if your company wants to move towards archetype 2:

- **Define a data strategy**: Think about possible data lacks and strategic decisions where additional data are needed. There should be a goal or expected target to reach that drives business decisions. The next step is to consider which additional data are needed to gain new insights to drive business decisions. These questions
7 Conclusion and recommendations
7.2 Research contribution

should be asked: (1) Which data do I have today? (2) Which data are collected and measured today? (3) Which other or new data are important? (4) What is the expected value or benefit of the new data?

- **Secure financial and operational feasibility:** After the decision to launch first Big Data efforts, think about investments in Big Data technologies: How do I finance this (new business models, lower costs)? What are my identified use cases?

- **Complement lack of knowledge with external help:** Identify knowledge gaps within your organization and if necessary contract 3rd parties to consult in certain knowledge areas.

- **Set a climate for organizational change:** The first step is to make sure that all employees understand the issue of Big Data, thus it firstly requires a small lesson about what Big Data is and which potential targets should be accomplished. Discuss the next steps together with your employees and be honest and open about ethical considerations and risks. Transparent communication is necessary to have all employees on board.

- **Build a Big Data platform:** There is need for new IT systems, which are capable of holistic data storage and more simplified IT architecture. A centralized data warehouse is beneficial. The Big Data platform should be incrementally scalable, modularized and flexible to grow with the number of use cases and added data sources. Customized and tailored to the specific needs to data structure and data sources of the company shall be prioritized over out-of-the-box solutions.

- **Facilitate coordination and cooperation across departments:** Implementing Big Data is a multi-skill task and requires the involvement of several departments. One department cannot tackle this issue alone. Open communication, manoeuvring company politics and soft change management skills are advised to enable coordination and cooperation across departments.

***if your company wants to move towards archetype 3:***

- **Negotiate and enable access to external data:** Not every business partner is best suited to exchange data with. It is recommended to prioritise partners with a long lasting business relation and where it makes sense in terms of volume to achieve substantial savings. Identifying mutual benefits for both data-sharing partners might increase the likelihood of data source access; the same applies to strengthening business partner ships. If one partner is not cooperating on data exchange, consider changing to a new one. How should these external data sources be evaluated? Calculate or estimate the trade-off between potential business value and the effort to gain access to these data sources as well as the technical requirements to integrate data in terms of financial investments and expected data quality.

- **Manage cultural change:** The view of data as a valuable asset must be anchored in the corporate culture. Additionally, there is need to trust data and overcome concerns with data privacy and cyber security. However, employees might be reluctant to change their views and working habits due to high risk aversion. Thus, it is important to consider and include all employees and engage them throughout the change process. Let them speak about their considerations and hesitations and solve them together. Their questions need to be answered about what happens, what are the risks and communicate all aspects in a transparent way to foster an open corporate culture. Thus, a step-by-step approach is advised, such as an evolutionary process.
7 Conclusion and recommendations
7.4 Limitations and further research

…if your company wants to move towards archetype 4:

- **Ensure sufficient supply of new digital talents:** Big Data requires a different skill set as work will get more complex due to increased automatization and digitization where IT-related skills are required. Target new digital talents from outside and make sure to lock-in younger talents through partnerships, scholarships or student work activities. Create an attractive workplace environment and invest in employer branding to be able to compete with start-up companies for the best analytical talent. Future employees must have the capability to understand Big Data and transfer it into business value. Moreover, provide education and training to constantly stay on top of new knowledge.

- **Mitigate risks constantly:** It is useful to think about future scenarios and prepare strategies to counteract them. Questions to ask are: How are current trends affecting my business strategy, how can my business model look like in the future, who is a potential disruptor and what are my exit options in case of failure? Furthermore, think about recovering the loss of critical competencies and knowledge from employees, when business operations are automatized and digitized. At last, do not lose focus from your core operations by focusing too much on Big Data and being innovative. It might be beneficial to have a balanced ambidextrous strategy, doing the operational work as usual but also being innovative and thinking about digitalization.

7.4 Limitations and further research

It has to be noted that the research findings have been acquired with thoroughly executed methodologies. However, the following research limitations have to be accounted for:

- Since several interviewees refused to be recorded transcription was not possible. Thus handwritten notes might not have captured all details, but following ethical research principles were considered more important.

- Not all interviews have been conducted in native language, thus it might lead to translation misunderstandings. Yet, English as a language is a widely used in workplaces, especially in international firms, and thus the implications are deemed minimal.

- All interviews were conducted via telephone, thus no additional non-verbal communication aspects such as body language or facial expressions could be observed. However, due to time and financial constraints, this research approach was nevertheless considered appropriate given the geographical reach.

- Data analysis was only carried out by a single author, hence no other reviewer was involved in the interview coding process, which might have resulted in biased interpretations and conclusions. However, whenever lack of clarity occurred internal and external supervisors were consulted to avoid misinterpretation.

- Difficulty to compare firms and draw universally applicable conclusions are present, due to different cultural context, different firm sizes and industries and interviewing only one employee representative from each company.

- One possible overly positive influence on the results of this research may be the fact that out of the large number of potential interviewees contacted those who eventually agreed to participate in this study might have been unusually highly interested in the research topic and therefore were more willing to share their (positive/negative) insights and knowledge.

In order to tackle these limitations, it will be beneficial to conduct a similar study with a more precise sampling strategy aiming to include similar companies in size, industry and geographical area. Moreover, a study targeting one particular industry and focused scope of
7 Conclusion and recommendations
7.4 Limitations and further research

One aspect of Big Data implementation might lead to more detailed and comparable findings. Regarding other research avenues, it is proposed to conduct a similar study collecting quantitative data to validate these qualitative research findings. Further validation of the Big Data archetype matrix might be beneficial, thus it is recommended to gather more empirical evidence to refine or criticize the matrix. Since the set of resources and capabilities related to Big Data is still evolving, another research avenue could be to conduct a longitudinal case study to investigate the implementation of Big Data over time or to perform a nested case study approach within one large organization.
8 Bibliography


Koch, V., Kuge, S., Geissbauer, R. & Schrauf, S., 2014. *Opportunities and Challenges of The Industrial Internet*, Available at: https://www.strategyand.pwc.com/de/studien/industry-4-0.


8 Bibliography


McKinsey Digital, 2016. Industry 4.0 after the initial hype - Where manufacturers are finding value and how they can best capture it, Available at: https://www.mckinsey.de/files/mckinsey_industry_40_2016.pdf.


8 Bibliography


PricewaterhouseCoopers (PwC), 2015. *Seizing the information advantage How organisations can unlock value and insight from the information they hold*, Available at: https://www.pwc.es/es/publicaciones/tecnologia/assets/Seizing-The-Information-Advantage.pdf.


8 Bibliography


INTERVIEW GUIDE

0 Before the interview

Thank you for taking your time to participate in this interview. For this interview, 30-60 minutes have been scheduled, is this still the appropriate time or do you need to leave earlier?

The aim of this interview is to gain a deeper understanding on data collection, data integration, Big Data analytics and implementation of Big Data projects. This interview is carried out in the scope of a Master's thesis project in the degree course "Industrial Management" at KTH Royal Institute of Technology Stockholm.

To obtain rich data and information from this interview, I will not talk that much but rather only ask questions and clarifications. Therefore, I just want to point out that I am not disinterested if I do not say much. I just do not want to lead you in any direction.

The participation is voluntary. At any point in time you can choose to discontinue the interview. All data will be treated as confidential and will be handled with highest care and diligence. The participation is anonymous and you will not be mentioned by name or any other features where it would be possible to draw a conclusion to your name, position or company.

To help me to analyse the interview material later, this interview will be recorded with your permission. We will start the recording and we will now repeat the question, is it ok for you to record this interview?

(Start recording the interview - official start of the interview)

1 Introduction – General questions about the participant

1. What is your position and role in the company?
2. What is your field of expertise?
3. For how long have you worked in your field of expertise?

2 Data collection and integration

2.1 General questions
1. Which types data are collected today?
2. Where are data lacks?
3. Where do new data sources come from?
4. From which areas can new data be collected?
5. What types do newly collected data have?
6. What is the difference between internal and external data?

2.2 Specific questions
1. How can multiple data sources be integrated?
2. Which approach do companies follow when integrating more data sources?
3. What needs to be considered when integrating more and more data sources?
4. What is the benefit of having multiple data sources?
3 Big Data Analytics

3.1 General questions
1. What is Big Data analytics?
2. What is the process of big data analytics?
3. Why is Big Data analytics needed?
4. What are the benefits of big data analytics?
5. What are the difficulties of big data analytics?
6. What is the value of big data analytics?

3.2 Specific questions
1. Are there challenges, and if so where along the process are the largest challenges to overcome?
2. Do the data need to be prepared and if so in what way?
3. Which different analytical methods are needed?
4. What happens if more data from different types are analysed?

4 Implementation services

4.1 General questions
1. Why do companies need implementation services?
2. Which kind of implementation services do you provide?
3. Where are the largest challenges in implementation for companies?

4.2 Specific questions
1. What are the technical challenges in implementation?
2. What are the organizational challenges in implementation?
3. What are the individual challenges in implementation?
4. Which new capabilities do need firms to develop in order to successfully implement big data analytics?
5. Which resources do firms need to acquire in order to successfully implement big data analytics?
6. Which resources or capabilities are not needed anymore or decline with the use of big data analytics? Why?

4.3 Very specific questions
1. How do companies transform into data-driven organizations?
2. What are the typical steps of such a transformation?
3. What are the dynamic capabilities needed to transform?

5 Closing
1. Do you have anything to add?
2. Do you have any last questions?
3. Can I send you follow-up questions if needed?
4. Do you know other people, who I can interview?

6 Wrap-up

Thank you again for taking the time to participate in this interview. The interview is now officially over and the recording will be stopped.

(Stop recording the interview - official end of the interview)