Gordon Institute of Business Science
University of Pretoria

# A practice for formulating an effective data-led strategy

Gavin Reuben

18361562

A research project submitted to the Gordon Institute of Business Science, University of Pretoria, in partial fulfilment of the requirements for the degree of Master of Business Administration.

**11 November 2019**

## ABSTRACT

This study is positioned to propose the enabling factors leading to the practice of formulating an effective data-led strategy (EDLS). Through the theoretical lenses of the resource-based view, absorptive capacity and attention-based view, from information systems theory, suitable first order constructs are proposed in relation to the second order constructs of resources, absorptive capacity and attention. These form antecedents to the higher order construct of EDLS.

Through a quantitative analysis, electronic surveys are used as a mechanism to collect data from managers within medium to large organisations, that have experience in both the elements of strategy and big data analytics. Leveraging a component analysis, reliability and validity testing, together with model assessment through the adoption of partial lease squares structural equation modelling, a structural model and associated measurement scale was developed, that has seven first order constructs which are related through the three second order constructs.

Not only has the study provided a theoretical model for the practice of achieving an EDLS but also enables managers in pursuit of data-led strategies to effectively manage resources, absorb relevant knowledge and focus their attention, aligned with their big data-related activities, to achieve the desired outcome of a competitive advantage.



**KEYWORDS:** big data, data-led strategy, resources, attention, absorptive capacity

## Declaration

I declare that this journal article and attached supplement is my own work. It is submitted in partial fulfilment of the requirements for the degree of Master of Business Administration at the Gordon Institute of Business Science, University of Pretoria. It has not been submitted before for any degree or examination in any other University. I further declare that I have obtained the necessary authorisation and consent to carry out this research.

_______________________ _______________________

**Gavin Reuben** **Date**

## Table of Contents

## COVER LETTER

### Motivation of journal choice

The journal chosen for this research article is Business Horizons which is Scopus indexed and has an Academic Journal Guide quality rating of 2 according to the March 2018 Association of Business Schools ranking. This research aims to bridge the gap between Information Systems and Management theory, focused on providing managers or strategy influencers a framework for developing a practice towards effective data-led strategy formulation. The type of research, the topic as well as the origin of the majority of referenced articles were all included as factors when choosing the journal.

The selected journal, Business Horizons is focused in the domain of General and Strategy and had a call for papers focused on digital transformation and big data related topics. Through communication with the editor as well as the factors listed, Business Horizons is the most appropriate journal for this article. While there is, a similar methodology proposed in supporting research, a capabilities view has been taken towards Big Data and the articles are focused in the realm of IT systems, Knowledge Management and Operations. Journals such as International Journal of Production Economics, Information and Management and Journal of Business Research were key resources to develop this research.

### Article details

The article was written in accordance with the Business Horizons journal guidelines, which was also confirmed with the editor of the journal, to eliminate any ambiguity. Attached in Appendix A is a copy of the guidelines together with the confirmation from the editor. The format of the article included a title page confirming the article title followed by the researcher details and supervisor details. In addition, a sample article, focused around the domain of Big Data, is included in Appendix B.

# 1. Literature review

## 1.1 Introduction

The chapter provides an overview by delving into the relevant literature related to the concepts, constructs and variables that underpin the proposed enabling factors that lead towards a practice of formulating an effective data-led strategy (EDLS). The aim of the research is to enable managers in medium to large organisations with the most effective organisational levers that could be used to exploit the value of big data (BD) and associated analytics capabilities. This is executed through the development of a conceptual framework and associated measurement scale, exploring and confirming relevant factors and associated measured variables.

To commence, the guiding philosophy is unpacked through a definition of strategy and the concepts relating BD and analytics. Through this definition, it is further required that BD is defined and characterised in terms of the differences from the traditional realm of data analytics. The value of BD and some of the challenges experienced by organisations are explored to understand the overall impact of leveraging BD capabilities to enable effective strategy. Through an understanding of the resource-based view of organisations, the absorptive capacity of new knowledge and how decision-makers pay attention to the business environment, the link between these constructs, associated latent variables and the higher order construct of EDLS, is proposed through a structural model. This leads to the development of the guiding research questions and formulated hypotheses that model the research.

## 1.2 The practice of strategy

In order to understand some of the requisites to enable a practice of EDLS, a view of strategy needs to be proposed.

Strategy has multiple definitions that differ based on the guiding philosophical school of thought underpinning the definition (Parnell & Lester, 2003). In one view by Casadesus-Masanell and Ricart (2009), strategy is defined as “the choice of business model, and the business model employed determines the tactics available to the firm to compete against, or cooperate with, other firms in the marketplace” (p.3). According to Mintzberg (1987) and Tawse, Patrick and Vera (2019), strategy formulation precedes action and should thus be deliberate in its purpose. However, it is also argued that intended strategy could result in both a deliberate and unrealised strategy (Mintzberg, 1987). This could lead to a strategy that is emergent and thereafter realised (Mintzberg, 1987). Given this, the notion is that a finalised strategy is not always a prerequisite to strategy implementation. Strategy can be realised through the process of formulation or implementation and be adapted to suit the desired outcome. This is further expanded by Martin (2014), that posits that the ability for strategy to be adaptive is a concept that has always existed, as the environment of business has always been dynamic and complex in nature.

Based on this, it is argued that strategic decision-making is a critical process that requires purposeful strategic choice with the ability to adapt based on the environment of business operation (Martin, 2014). The value of BD and the ability to achieve a competitive advantage by effectively leveraging BD capabilities, is widely understood and accepted by organisations as a non-negotiable strategy to employ (Srivastava & Gopalkrishnan, 2015).

### 1.3 The value of Big Data

Prior to articulation of the potential value from BD, a view on what defines BD is provided. The most common view defines BD as the collection of large amounts of data that requires specialised computing platforms and skills to analyse the data, based on the size of the data sets (Chen & Zhang, 2014). According to Chen and Zhang (2014), the ability to derive insights from this data is critical yet challenging. There are a host of prerequisites that organisations require such as capabilities and operations, to support BD strategies, before value can be derived (Mazzei & Noble, 2017). According to Lee (2017) and Chen and Zhang (2014), the five defining characteristics to qualify BD are volume, velocity, variety, veracity and value. These characteristics are summarised as follows:

*Volume* refers to the quantity of data available for processing, which traditional database platforms are unable to support based on the scale of the data sets (Lee, 2017; Chen & Zhang, 2014).

*Velocity* refers to how quickly data is generated and required to be processed, which in some instances requires real-time availability and analysis to generate insights for decision-making (Lee, 2017; Chen & Zhang, 2014).

*Variety* is based on the various types of data and sources that could be regarded as being both structured and unstructured in nature, where examples of unstructured data is voice, video and text (Lee, 2017).

*Veracity* links to the inaccuracy and ambiguity of data that results in inaccuracy of insights hence leading to a low adoption in organisations (Lee, 2017).

*Value* is based on the competitive advantage that can be created for organisations (Lee, 2017). This value is created through insights gained, to enable effective strategic decision-making (Chen & Zhang, 2014).

While these are posited to be the five main characteristics for the definition of BD, this list is not exhaustive. Further insights from Lee (2017), mentions the expansion of these characteristics to include complexity and variation, however these could be combined in the definitions presented by velocity and variety.

In the context of value, access to the large amounts of high integrity customer data is the foundational layer in building a data-led strategy (Mazzei & Noble, 2017). Understanding the definition of customer value and how to engineer this by leveraging BD capabilities, becomes the next critical step in becoming data-led (Mazzei & Noble, 2017).

Using this data to innovate is the evolution that organisations must adopt to remain relevant (Fink, Reeves, Palma, & Farr, 2017; Mikalef, Boura, Lekakos, & Krogstie, 2019). The key insights gained through access to large amounts of conformed quality data, holds the promise of the ability to enable insights that lead to dynamic decision-making and the ability to increase the productivity within organisations (Mcabee, Landis, & Burke, 2017). Further to this, organisations operate in an environment of business that is riddled with complexity, variability, uncertainty and ambiguity and therefore there is a need for strategy to be adaptive in these complex environments (Martin, 2014). This ability to be adaptive is critical, and the capability of having dynamic data-driven insights is mentioned to be a primary driver to enable organisations to be agile in their decision-making process when adapting strategy (Kitchens, Dobolyi, Li, & Abbasi, 2018).

### 1.4 Challenges in building an EDLS

The concept of leveraging data to enlighten decision-making is one that organisations have been accustomed to (Mazzei & Noble, 2017). Through the evolution of the environment we live and operate in, the volumes and types of data available have also evolved (Wamba et al., 2017). The ability to derive insights from this scale and form of data requires a shift in technology, capabilities, resources, culture as well organisational focus and capacity, to enable effective value creation for organisations (Carruthers & Jackson, 2018).

Based on this, organisations are required to build new capabilities and require both technical and non-technical competencies that may not already be available to them (Wang, Xu, Fujita, & Liu, 2016).

At a technical level, skills and IT infrastructure is required to manage the large complex structured and unstructured nature of data and develop the ability to generate insights from this, for strategic decision-makers (Wang et al., 2016). Further to this, if organisations have already been in operation, they have existing IT infrastructure and systems that need to evolve and interact with BD enabled platforms and systems, to effectively build the capability to use data as a tool (Mazzei & Noble, 2017; Wang et al., 2016).

Wang et al. (2016), further postulates that organisations are also faced with non-technical challenges that arise when there is an intention of building out a data-led strategy and BD capability. These challenges relate to leadership focus on driving organisational transformation and effectively driving a change in culture that is needed to effectively build out a data capability (Carruthers & Jackson, 2018). In addition, the element of resources becomes critical in developing a capability to enable strategy. According to Mazzei and Noble (2017), the old paradigm of strategy driving the data requirements in an organisation has shifted to data-driven insights driving strategy formulation and implementation. Further to this, the roadmap for building out a data capability within an organisation can happen at various value levels. Organisations need to be very specific on their strategic intent for their data capability, to assess the extent of their organisational transformation required. These levels are defined by data being a tool to enable decision-making, data creating new industries or forming pure data-led organisations (Mazzei & Noble, 2017). Each of these levels requires varying levels of transformation. At a data-led strategy level, there is a critical reliance on the impact of leadership in driving change of culture and engaging the right level of skills and resources to execute on a BD enabled transformation (Carruthers & Jackson, 2018; Mazzei & Noble, 2017). Therefore, managers need to be armed with the relevant knowledge that allows them to effectively manipulate the correct levers within an organisation, to reap the promise of an EDLS.

### 1.5 Resource-based view

The theory of the resource-based view (RBV) of a firm postulates that organisations are able to gain a competitive advantage from a strategic lens, through in-depth understanding, strategic placement and utilisation of resources, throughout an organisation (Barney, 1991; Raguseo & Vitari, 2018). Through active management of the resources, decision-makers and influencers to organisational strategy are able to drive the desired outcome (Akter, Wamba, Gunasekaran, Dubey, & Childe, 2016; Barney, 1991; Raguseo & Vitari, 2018).

The RBV theory postulated by Barney (1991), which is further expanded through academic research by Raguseo and Vitari (2018) and Rivard, Raymond and Verreault (2006), considers the diverse nature of resources and immobility through the understanding of four key characteristics:

- *Value*: This is defined as the ability to enable effectiveness through specific resources.
- *Rarity:* This factor considers the shortage of resources in comparison to competitors.
- *Imperfectly imitability:* Suggests that other organisations cannot easily gain access to these resources that are considered valuable and uncommon.
- *Non-substitutable:* These are resources that do not have alternate options to enable a competitive advantage.

Given this view on resources, BD related enabling factors are considered through the RBV theoretical lens, to determine the influence towards developing a practice for formulation of an EDLS. These factors are posited to form precursors to the construct of resources. Through the theoretical lens of RBV, the factors are assessed to determine their level of influence towards the overall management practice in formulation of an EDLS. In order to measure the impacts of resources on EDLS, big data as an asset, data infrastructure, technical skills of employees in relation to BD and data culture, are considered as factors for evaluation (Dubey, Gunasekaran, Childe, Blome, & Papadopoulos, 2019; Ghasemaghaei, 2019; Gupta & George, 2016; Wamba et al., 2017).

**Big Data**

Data is considered a fundamental resource and asset, required for enabling a data-led strategy (Ghasemaghaei, 2019; Ghasemaghaei, Hassanein, & Turel, 2017;

Gupta & George, 2016). As described in scholarly literature, regarding the concept of BD, both structured and unstructured data, from internal and external sources, lead to the formulation of a strategic data asset, within an organisation (Chen & Zhang, 2014; Gupta & George, 2016; Lee, 2017; Mazzei & Noble, 2017). The value of this asset is a result of access to large data sets, a variety of data, in high volumes that has high accuracy, to enable effective value creation (Ghasemaghaei, 2019; Mazzei & Noble, 2017). Based on this, big data as an asset is explored, through the lens of RBV, to understand the impact on strategy in a BD enabled environment. Given the view proposed by the RBV, big data as an asset should be unique to the organisation and have the ability to provide value to its customers that is unable to be supplied by competitors.

**Data Infrastructure**

As suggested by Wang et al. (2016), data infrastructure in the form of platforms and systems is an example of a critical resource in the domain of managing BD elements. According to McAfee and Brynjolfsson (2012) and Wamba et al. (2017), technology in the form IT infrastructure, specifically data infrastructure, is a critical resource that requires integration into the organisations IT landscape, to manage data assets and enable value creation. Further to this, IT infrastructure within an organisation is positioned to be one of the necessary factors in the construct of structural readiness, to enable value creation through big data analytics (BDA) (Ghasemaghaei, 2019). Through assessment of the technological readiness in the context of BD and the ability to integrate with legacy infrastructure, data infrastructure is assessed as an enabling factor, through the lens of RBV, to enable strategy (Ghasemaghaei, 2019; Gupta & George, 2016; Raguseo & Vitari, 2018; Wang et al., 2016). This infrastructure becomes critical in processing big data assets to derive the required information that is suggested to lead to insights thereafter wisdom (Adrian, Abdullah, Atan, & Jusoh, 2018).

**Technical Skills**

Skilled data personnel resources such as data analysts, data scientists and data engineers are still rare resources that are required to extract, transform and load data, thereafter derive relevant value through advanced analytical techniques, to enable business strategy (Ghasemaghaei, 2019; McAfee & Brynjolfsson, 2012). Given the recent formulation of the capabilities associated with BD, the appropriate experience, education and technical skills of human resources, become a critical component when developing a big data analytics capability (BDAC) (Akter et al., 2016; Chen & Zhang, 2014; Gupta & George, 2016; Wamba et al., 2017). According to McAfee & Brynjolfsson (2012) and Wixom, Yen, & Relich (2013), the enablement of strategy for an organisation is reliant on the effective utilisation of its personnel to execute on its strategic objectives. While the skills are transferable, the ability to completely be effective as a technical resource requires the tools in terms of infrastructure and unique data, to effectively develop value (Wixom et al., 2013).

**Data Culture**

Organisational culture is often referred to as a factor that impacts the way organisations respond to both internal and external changes in their environment, and how this enables their strategic objectives (Dubey, Gunasekaran, Childe, Roubaud, et al., 2019). Given this definition, there is a tacit link between the culture of an organisation and the strategy. Furthermore, Schein (2004) , defines culture as "a pattern of shared basic assumptions that was learned by a group as it solved its problems of external adaptation and internal integration, that has worked well enough to be considered valid and, therefore, to be taught to new members as the correct way to perceive, think, and feel in relation to those problems" (p.17). Organisational culture can therefore be defined as the these shared assumptions, values and beliefs, that shape and define the way that organisations respond to changes (Dubey, Gunasekaran, Childe, Blome, et al., 2019; Dubey, Gunasekaran, Childe, Roubaud, et al., 2019; Rude, 2014; Schein, 2004). As described by Carruthers and Jackson (2018), the ability to drive a data-led strategy, requires transformation and culture is one of the critical factors towards leading this change. Given this requirement, defining and understanding the elements of a data-driven culture is explored as an enabling factor for an EDLS. Further to this, data culture is a resource that organisations must possess to extract maximum value from their data strategy, to enable a competitive advantage (Gupta & George, 2016; Frisk & Bannister, 2017).

### 1.6 Absorptive capacity

Innovation is defined as the ability to find a new idea, method or product and can either be radical or incremental in nature (Mikalef et al., 2019). This ability is posited to be as a result of the BDAC within an organisation, resulting in value creation through data-led strategy (Chen, Lin, & Chang, 2009; Mikalef et al., 2019; Solís-Molina, Hernández-Espallardo, & Rodríguez-Orejuela, 2018).

Absorptive capacity theory (ACAPT) was developed through seminal work by Cohen and Levinthal (1990), that postulates that an organisations' ability to absorb new information and effectively build a capability through this, is imperative for value creation. To derive an effective strategy and develop an ability to be innovative, the capacity available in the organisation to continually learn from the environment and develop competencies and capabilities is defined as the organisations absorptive capacity (Cohen & Levinthal, 1990; Solís-Molina et al., 2018). Based on this, the absorptive capacity of an organisation could be postulated to be a necessary factor, for building an effective BDAC, to enable dynamic decision-making and adaptive strategy (Solís-Molina et al., 2018).

This concept is further solidified through research by Rodriguez and Da Cunha (2018) and Wamba et al. (2017), that suggest that absorptive capacity is seen as an enabler to BDAC. Absorptive capacity exists at both the organisation and employee level however, it is stated that while organisational absorptive capacity will develop cumulatively based on the individuals' levels, there are also distinct differences between them both (Cohen & Levinthal, 1990). Being based on individuals at some level, the overall organisational absorptive capacity is therefore impacted by the levels of absorptive capacity available in employees. These traces of absorptive capacity from employees are situated between the organisation and external environment as well as between various divisions in the organisation (Cohen & Levinthal, 1990; Roberts, Galluch, Dinger, & Grover, 2012).

Due to the fact that the lens of ACAPT is considered in this context, there are necessary factors that are required for consideration (Roberts et al., 2012). As it is posited that absorptive capacity can influence EDLS, the ability for knowledge acquisition, assimilation, transformation and exploitation are relevant factors that form the process of absorption, and will be validated (Cohen & Levinthal, 1990; Rodriguez & Da Cunha, 2018; Xie, Zou, & Qi, 2018).

### Acquisition

Acquisition is defined by the ability to learn, attain or develop knowledge(Xie et al., 2018). Further to this acquisition considers how this newly found knowledge can be leveraged to drive or enable strategic objective execution (Xie et al., 2018).

In the context of knowledge acquisition, the organisational absorptive capacity is posited to be impacted by the direction, speed and intensity of the effort (Zahra &

George, 2002). Based on this, knowledge acquisition should be focused on the requirements of the organisation and is therefore, a measure of the ability to identify, value and obtain knowledge external to the organisation (Camisón & Forés, 2010; Cohen & Levinthal, 1990; Xie et al., 2018).

**Assimilation**

Knowledge assimilation delves into the available capacity of an organisation to obtain external information and fully understand this information or ideas, utilising it within their organisation (Camisón & Forés, 2010; Zahra & George, 2002). In the context of BD and advanced analytics, there is an opportunity not only to consume data assets from external sources but the ability to obtain external knowledge and through the process of deep understanding, leverage this knowledge to fit their organisational understanding (Xie et al., 2018). Assimilation thus requires available capacity in the organisation to effective leverage its acquired knowledge.

**Transformation**

Transformation is suggested to delve into the capabilities within an organisation that develop or create a new process to combine the new knowledge attained with existing knowledge, leading to a strategic change (Cohen & Levinthal, 1990; Zahra & George, 2002). The ability to transform relies on the organisations capacity to add new, remove old or combine new and old knowledge, to promote innovation (Camisón & Forés, 2010; Xie et al., 2018). Further to this, Xie et al. (2018), also proposes that the ability to transform knowledge that is gained creates the capacity for more new knowledge to be absorbed into the organisation. Given these conditions, transformation is a necessary practice to promote value creation (Yli-Renko, Autio, & Sapienza, 2001).

**Exploitation**

The exploitation of external knowledge, according to (Cohen & Levinthal, 1990), is a mechanism that is critical for organisations to develop innovative capabilities. Exploitation, in the context of ACAPT, considers the organisations capacity to leverage the external knowledge gained to create the opportunity for a competitive advantage (Camisón & Forés, 2010; Xie et al., 2018). Given this definition, exploitation in the context of EDLS is concerned with ensuring that maximum value is extracted from knowledge gained in the realm of BD.

## 1.7 Attention-based view

Attention is a cognitive process of choosing to focus on certain information while making the trade-off of what information to ignore (Ocasio, 1997). In the context of organisations, the focus attention becomes a critical choice that needs to be driven by decision-makers, to propel the organisation in the intended strategic direction (Palmié, Lingens, & Gassmann, 2016). According to Palmié et al., (2016), these organisational choices are a result of the actions from decision-makers and evidence of where attention has been focused.

According to the attention-based view (ABV), organisations manage attention through three related concepts of how attention is focused, the situation surrounding the direction of the attention and the organisational structural influence on the distribution of attention (Ocasio, 1997). The basic premise is that focus of attention is driven by situational factors of the decision-makers and that these situational factors are a result of organisational structural and operational choices, in relation to the positioning of decision-makers (Ocasio, 1997). It is further noted that attention is finite in nature and therefore the choice of where to focus is crucial to drive the allocation of organisational resources and that the channels for communication within an organisation and externally, promote the adaptability of strategy (Ocasio, Laamanen, & Vaara, 2018).

Influences of an attention-based view of organisations in building an EDLS, is considered through the factors of where attention is focused, the situation surrounding attention as well as the organisational structures and channels that influence attention (Gebauer, 2009; Ocasio, 1997; Ocasio et al., 2018; Palmié et al., 2016). Therefore, these factors are subjected to validation.

### Focus

Within organisations, the focus of decision-makers or leaders is directed towards the areas they want to influence and the spaces in which they execute (Gebauer, 2009; Ocasio et al., 2018). Given the limited capacity that decisions makers have within organisations and the plethora of areas that require attention, the decision of where to focus attention needs to be aligned to the strategic direction of the organisation (Joseph & Wilson, 2018; Ocasio, 1997; Ocasio & Joseph, 2005). Further to this, the actions that result from decision-makers are an indication of the issues and answers that focus has been directed towards (Gebauer, 2009).

### Situational

Situated attention posits that decision-makers and leaders attention is based on the situational context they find themselves in (Ocasio, 1997). Based on this, attention can change given adapting situational environments (Ocasio et al., 2018). Given this theoretical view on situated attention, Ocasio (1997) further unpacks that varying situations alter the focus of decision-makers, thus leading to a channelling of attention given the situational context. Based on this potential shift of focus, the actions of decision-makers will be as a result of the situational context (Gebauer, 2009).

### Structures and Channels

Further to the focus of attention and the situational context, the organisational structures and channels are proposed to impact how attention is dispersed across the organisation (Gebauer, 2009; Ocasio, 1997; Ocasio & Joseph, 2005; Ocasio et al., 2018). The distribution of attention has multiple factors that should be considered to understand the impact which includes and are not limited to the organisational rules, resources available, allocation of decision-makers between organisation functions, ability to communicate and the social context of the organisation (Gebauer, 2009; Ferreira, 2017). According to Ocasio (1997), every function within an organisation is comprised of relevant communication channels and procedures that govern the operation. Given this factor in addition to the potential situation exposure, the focus of decision-makers attention is channelled by the governing structures (Joseph & Ocasio, 2012; Palmié et al., 2016). Palmié et al., (2016), further posit that the structural channels are comprised of both physical and electronic mediums that are related to communication and procedural channels.

## 1.8 A proposed conceptual research model

Given the need for organisations to develop and understand the practice of enabling an EDLS, the research aimed to explore and develop a proposed conceptual model through prediction, leading to theory development. Based on the literary overview, it was predicted that resources, absorptive capacity and attention in an organisation, are enabling factors to develop this practice. As described, the theoretical lenses of RBV, ACAPT and ABV are applied to develop related latent variables that link to these constructs, through the context of BD and advanced analytics. This is postulated to lead to the benefit of value creation and a competitive advantage given the challenging environment of business, that organisations find themselves in (Raisch, Birkinshaw, Probst, & Tushman, 2009). The aims of the research can therefore be summarised according to the following three research questions:

*$R_1$: Is there an influence of data Resources on EDLS*

*$R_2$: Is there is an influence of organisational Absorptive Capacity on EDLS*

*$R_3$: Is there is an influence from organisational Attention on BDAC on EDLS*

Considering a RBV of an organisation, the positive impact of BD resources enabling a competitive advantage, is suggested as a result of the latent variables big data assets, data infrastructure, the skills of employees in the BD team and organisational culture (Akter et al., 2016; Barney, 1991; Ghasemaghaei, 2019; Gupta & George, 2016). The premise of RBV suggests that an effective strategy, that is data-led can be informed through the integration of the specific resources suggested, as long as they meet the criteria of being uncommon, imperfectly imitable, of value and not able to be substituted (Barney, 1991; Rivard et al., 2006). Based on the above, the following hypothesis was proposed:

*$H_1$: There is a significant positive influence of Resources on EDLS*

The theoretical overview on the absorptive capacity of an organisation, suggests that the ability for organisations to absorb new information and effectively leverage this leads to the potential of a competitive advantage (Roberts et al., 2012). Mazzei and Noble (2017), provided an evolutionary data view which suggests that data can be interpreted to form information, leading to insights and thereafter wisdom, through organisational capability and culture development. Based on this, the ability to attain or acquire, assimilate, transform and exploit information in the BD landscape, leads to the opportunity of a competitive advantage through a successful strategy (Roberts et al., 2012; Xie et al., 2018). Based on the above, the following hypothesis was proposed:

*$H_2$: There is a significant positive influence of Absorptive Capacity on EDLS*

At an organisational level, the attention given towards BD and the potential influence of this capability towards an effective strategy is suggested to be as a result of multiple latent variables (Ocasio & Joseph, 2018). This attention, at organisational level, is composed of attention of individual decision-makers within an organisation, and therefore the factors of focus, situational as well as organisational structures and channels, are suggested to result in an effective strategy given organisational attention towards BDAC (Gebauer, 2009; Joseph & Ocasio, 2012; Ocasio et al., 2018; Palmié et al., 2016). Finally, the last hypothesis suggested was:

*$H_3$: There is a significant positive influence of Attention on EDLS*

To conceptualise this study, a model for an EDLS is proposed through a multidimensional third-order latent construct, through influences of the latent second order constructs of resources, absorptive capacity and attention (inner model) (Chang, Franke, & Lee, 2016; Hair, Ringle, & Sarstedt, 2011; Hardin, Chang, & Fuller, 2008). Further development of this framework is conceptualised through eleven first order sub-dimensions with respective measurement models (outer model) through a questionnaire detailed in Appendix C (Hair et al., 2011). The model proposed is shown in Figure 1 below with reflective indicators.

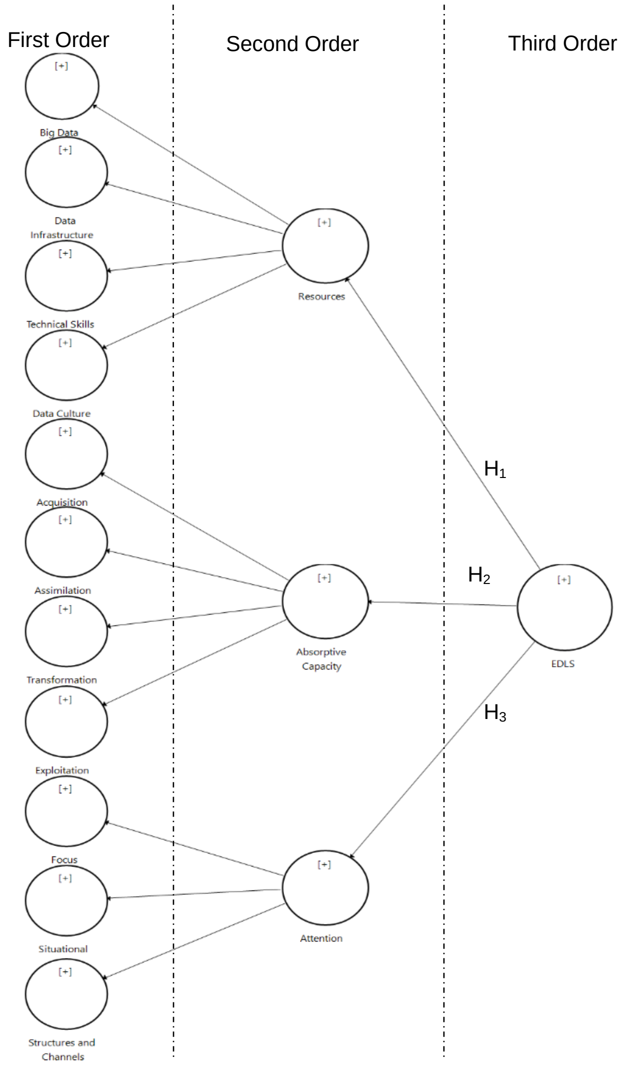



*Figure 1: EDLS conceptual model*

## 1.9 Conclusion

The literature provided, delved into the related constructs and representative first order sub dimensions, that lead to the development of a proposed conceptual model. Through measurement of the first order sub dimensions and related theoretical links of strategy and BD, the influence of suggested factors were assessed. This adds to the body of knowledge surrounding information systems literature and management theory, in the context of organisations. Through the lenses of RBV, ACAPT and ABV, the suggested first order sub dimensions are illustrated in Figure 1. This is a proposed alternate view on BD, developed from the research and design of the conceptual model on BDAC, to enable firm performance (Akter et al., 2016; Gupta & George, 2016; Wamba et al., 2017). This perspective focused on the management view of BD enabled strategy, therefore the research provided businesses in the quest for transformation, a strategic management framework.

## 2. Proposed research methodology and design

### 2.1 Introduction

The research aim was centred on delivering a practice for managers within organisations, with intentions of leveraging BD as a source of competitive advantage. This was executed on by defining the enabling factors that lead to a practice, for the formulation of an EDLS, through a theoretically posited model defined in Figure 1. To deliver on this requirement, the researcher developed a structural model that was tested for appropriate relevance, together with an associated measurement scale. This methodology leveraged a process of component analysis, reliability and validity testing together with a research approach grounded in Partial Least Squares Structural Equation Modelling (PLS-SEM).

### 2.2 Choice of methodology

The philosophy for the research was based on a positivism view given the researchers orientation together with the influence from similar information systems based studies (Abbasi, Sarker, & Chiang, 2016; Gemma, 2018). The methodology for the research, therefore employed a deductive approach, to test for influences. This was conducted between the selected second order constructs of resources, absorptive capacity and attention, with their respective factors through first order constructs, against the higher order construct of EDLS. This was undertaken through a quantitative analysis, based on responses of the sample being surveyed. Based on the philosophy of the study, it was possible to generalise the results from the sample to the population, provided that the sample was able to represent the population in scope for the study (Creswell, 2014; Zikmund, Babin, Carr, & Griffin, 2009).

Based on the positivist view, the survey responses allowed for theory testing through a deductive approach (Creswell, 2014; Saunders & Lewis, 2012). The research aimed to gather specific data based on the relationships to organisational strategy and the link posited between organisational strategy and data-led strategy. This was in relation to all constructs and respective variables. This allowed for testing the rigour of the proposed conceptual model.

A mono-method approach was adopted as a single data collection technique via an electronic channel (Creswell, 2014). These responses were collected through this channel, given the ease and efficiency of distribution and completion, for the respondents. The study employed a quantitative research method that allowed for testing of the theoretical research proposed through observation. This included testing the validity of the conceptual model presented in Figure 1, that postulates that the influences of resources, absorptive capacity and attention are positive and significant in formulating a practice, leading to an EDLS.

The research design was deemed explanatory as its purpose was to discover and quantify influences on the ability to define and execute on an EDLS. This was achieved through the consideration of the organisations ability to manage the optimal mix of resources to enable value creation, absorb new information from the environment and the ability to direct the organisational attention, through its decision-makers (Barney, 1991; Cohen & Levinthal, 1990; Ocasio, 1997; (Zikmund et al., 2009). While the theories of RBV, ACAPT and ABV are able to address innovation and value creation through effective strategy, there has been little research available that looked into the impacts of these views on organisations that intend on executing a data-led strategy.

The research strategy was executed through the survey distribution with all relevant stakeholders within medium to large organisations (with over 50 employees) that have BDAC and intentions of enabling data-led strategies (de Wet, 2019). This criterion on organisational size is based on the premise that medium to large organisations varies from small organisations in the context of strategy formulation and development when considered through a RBV lens (Grant, 1991). Further to this, the impact of BD in the context of this research assumed that the process of surveying was designed to test the impacts of the main constructs, relative to EDLS (Zikmund et al., 2009). In order to statistically test the relevance of these constructs, the data was gathered, analysed and discussed, to provide insights on the relevance of the proposed conceptual model (Zikmund et al., 2009).

Given the low maturity in this field of study and limited data, a cross-sectional based time frame was used to conduct the research. The process of surveys was used to understand the current views surrounding the second order constructs through respective first order constructs, in relation to the effectiveness of a data-led strategy, therefore warranting the need for a cross-sectional view (Creswell, 2014). This was seen to be largely based on the maturity of data capabilities in the organisations represented through the sample.

### 2.3 Population

Based on the aim of the research and associated objectives, the population under research were respondents in all organisations that have the intention of a data-led strategy and have BDA capabilities to enable value creation. To ensure that respondents were within this population, control questions were used to assess the applicability of the study to relevant organisations. Surveys were also purposefully sent to relevant groups of individuals given the researchers' personal networks.

The respondents targeted were ideally senior or middle management employees, who were involved in, or were collaborators to, the BDAC within the organisation. It was crucial to include employees that are decision-makers regarding data strategy, to get a sense of how their attention is focused (Ocasio et al., 2018). Further to this, absorptive capacity required consideration at both the individual and organisational level (Soo, Wei Tian, Teo, & Cordery, 2017). Therefore, all relevant stakeholders that were involved with usage or development of the BDAC, were targeted. While a specific industry or country should have been contained within the study, the lack of participant scale warranted multiple industries and countries to be included.

### 2.4 Unit of analysis

The unit of analysis was defined to be each individual within organisations that have the intention and experience with data-led strategies and BDAC to support this (Zikmund et al., 2009). These individuals were considered across multiple industries and organisations to get a representative view the population. The premise was that these individuals were involved in the use or creation of BDAC and insights to deliver value in their respective organisations. It was also established if the individual was aware of the impacts of BD and analytics on strategy.

### 2.5 Sampling method and size

The quantitative nature of the research added the necessity of required responses from a large enough sample size to ensure reliable and valid data points are collected, enabling beneficial results through analysis (Creswell, 2014; Hair, Black, Babin, & Anderson, 2010). The researcher was required to test theoretical appropriateness through deduction, therefore there was a need for a representative sample, that also allowed for the generalisation from of the results (Zikmund et al., 2009). While the more samples available, the more representative the results, it was impractical to obtain an excessive scale given access to respondents via personal networks and the available time (Creswell, 2014).

Given the context of BDA and associated capabilities within the broader market and industry, a list of the population was not likely to be obtainable. As a result, probability sampling was not an applicable technique (Creswell, 2014). Based on this, a non-probability technique was explored (Zikmund et al., 2009). The data required for the research was based on knowledge, exposure and experience held by employees in organisations with BDAC and intentions of embracing data-led strategies. To extract this data and gain relative insights, surveys were distributed specifically to individuals in these types of organisations. This implies that a purposive sampling technique was utilised (Creswell, 2014).

The minimum sample size for this research was calculated based on the minimum requirement suggested by the proposed statistical technique being adopted (PLS-SEM). Hair et al. (2010), proposes a sample size based on a rule of ten times the number of links that point towards a latent variable in a structural model. However, Hoyle (1995) advises that even though PLS-SEM techniques are robust enough to handle small sample sizes, a sample size of between 100 to 200 is suggested by prior research in path modelling. In addition, the nature of the research warranted the requirement for a subset of the sample to conduct an Exploratory Factor Analysis (EFA), while a separate sample was used to conduct the relevant statistical model fit test. The researcher thus aimed to obtain a minimum sample size of 30 to 100 samples for the EFA and utilised a Principal Component Analysis (PCA), to do so (Carpenter, 2018; Johanson & Brooks, 2010). In addition, the decrease in sample size was accounted for based on data validation and removal of incomplete data.

### 2.6 Measurement instrument

According to guidance prescribed by Creswell (2014) and Zikmund et al. (2009), quantitative explanatory research is best supported through the use of surveys and extensive academic literature with related theoretical constructs. This provides a view of previous relationships tested through deductive research and provides an opportunity to define a new conceptual model through quantification of relationships between relevant constructs and associated factors.

While the questionnaires were used as the most appropriate tool for data collection, the use of a five-point Likert scale together with relevant questions was used to test path linkages between the measured variables and constructs under analysis. This was used to ensure consistency in the measurements across all constructs under study (Zikmund et al., 2009). This questionnaire was developed through leveraging and adopting existing measurement scales. These contained questions relating each of the measured variables proposed for the first order constructs in relation to the second order constructs of resources, absorptive capacity and attention.

The researcher had allocated ample time to receive an appropriate sample size as calculated above. This was estimated to be five weeks however, this timeframe was manipulated based on the number of valid samples received and prescribed by the research methodology. Only data relevant to the research was collected from the surveys and control questions were used to validate the population being targeted. Based on this, irrelevant and incomplete samples were removed. The full questionnaire utilised in the study is provided in Appendix C. The questionnaire was divided into three sections after the controlling and demographic questions were assessed. These sections were representative of each of the second order constructs in the proposed model. All scales were adapted based on the relevance to data-led strategy and respondents were asked to frame their response with this context in mind.

When assessing the construct of resources, the first order constructs of data infrastructure, technical skills, big data assets and data culture are assessed with four related measurement items per a first order construct (Dubey, Gunasekaran, Childe, Roubaud, et al., 2019; Ghasemaghaei, 2019; Gupta & George, 2016; Mikalef et al., 2019;  Wamba et al., 2017).

Secondly, the construct of absorptive capacity is posited to be related to the first order constructs of acquisition, assimilation, transformation and exploitation (Camisón & Forés, 2010; Xie et al., 2018; Zahra & George, 2002). These scales included four related measurement items per a first order construct assessed.

Finally, the construct of attention is related to the three first order constructs of focus, situational as well as structures and channels. These were also assessed with four related measurement items per a first order construct (Brattström, Frishammar, Richtnér, & Pflueger, 2018; Gebauer, 2009; Joseph & Ocasio, 2012; Palmié et al., 2016).

### 2.7 Data gathering process

Based on the quantitative nature of the study, the researcher purposefully distributed surveys to relevant stakeholders at various organisations through leveraging personal networks given personal exposure to BDAC, across industries. In addition, the Big Data and Analytics group on LinkedIn was used as a platform to target applicable respondents.

SurveyMonkey® was used as the online electronic tool that provided easy to use, convenient access to respondents. It also allowed them to undertake the survey in their own time, in a conducive environment. Only primary sources of data were considered through the survey responses, given the low maturity in existing similar models and scale development research, on BD related topics.

### 2.8 Analysis approach

The multivariate statistical method of PLS was adopted in this research to assess the research hypotheses. The data was first assessed by understanding the extent of incomplete data within the final total sample of 235 responses. Newman (2014) states that missing data within a survey design can be attributed to item level, construct level and person-level missingness.

Due to the confidentiality of the responses, each level of missingness could not be tested, and the researcher, therefore assumed a person level missingness based on the premise of missing at random (MAR) (Newman, 2014). The missing data was imputed based on the methods described by Hair et al. (2010) and (Scheffer, 2002). Based on this, any respondents with less than a 50% completion rate, was excluded from the research. All missing data from the remaining respondents, with less than 100% completion was imputed using the average of the industry in which they were currently employed. The final sample size for the research was N=107 which was deemed adequate for the research design. In addition to this, a sample of N=54 was used to conduct the PCA.

The PLS method forms part of the SEM family, which has increasingly become popular within empirical research (Rigdon, Sarstedt, & Ringle, 2017). This second generational technique provides a robust statistical technique that combines the statistical methodologies of regression and factor analysis (Hair et al., 2010). In addition, the PLS method does not require a large sample size, is recommended for both formative and reflective models and is efficient at parameter estimation (Hair, Sarstedt, Hopkins, & Kuppelwieser, 2014). Smart PLS 3.0 was adopted to assess the research model, which is also adopted by previous research within the BD domain (Akter et al., 2016). Based on the recommendations of Chin (2010) and Hair Jr. et al. (2017), the PLS algorithm, bootstrapping and blindfolding methods, within the Smart PLS software, was adopted to assess the inner model (structural), outer model (measurement), significance and model fit respectively. As depicted in Figure 2, a multi-stage process was adopted to assess the hypothesised research model (Hair, Hult, Ringle, & Sarstedt, 2016).

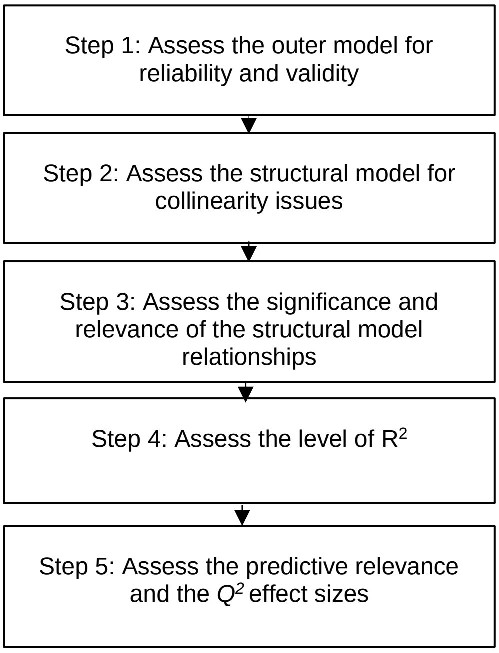



*Figure 2: PLS model evaluation process. Adapted from "A primer on partial least squares structural equation modelling (PLS-SEM)," by J.A. Hair, G.T.M. Hult, C.M. Ringle and M. Sarstedt, 2016, Sage, edition 2, p. 202. Copyright 2017 J.A. Hair, G.T.M. Hult, C.M. Ringle and M. Sarstedt.*

Following the methodology described by Hair et al. (2017), the assumptions for the PLS model was assessed, which included a PCA. Since the hypothesised research model had not been tested before, to the knowledge of the researcher based on research scanning, it was imperative to validate the compositions of the hypothesised constructs. The PCA technique is useful in dimension reduction and grouping measured variables based on the common variation (Koryak, Lockett, Hayton, Nicolaou, & Mole, 2018). Braeken & Van Assen (2017) posit that in order to effectively determine the factors that should be retained, the Kaiser's Eigenvalue one criterion must be adopted. This means that any factors with less than an eigenvalue of one, were not relevant and considered in the determination of the construct groupings.

To improve the validation and interpretability of the PCA analysis, rotational methods were also adopted. By including the Varimax rotational method, the maximization and minimization of high and low component loadings ensured a degree of construct determination accuracy (Costello & Osborne, 2005). Following the recommendations by Hair et al. (2010), Bartlett's test for sphericity and the Kaiser Meyer Olkin (KMO) values were also interpreted. KMO provides a measure of the sampling adequacy within the composed measured variables and requires a measure that exceeds the threshold of 0.5 at minimum for adequacy, where a value closer to 1.0 indicates an increased the level of adequacy (Pallant, 2007). The Bartlett's test for sphericity, on the other hand, compares the measured variables within the hypothesised group against correlations with its identity matrix. PCA can only be assessed if this test displays a significance of $p < 0.05$ (Hair et al., 2010). The pilot test sample of 54 respondents was adopted by the researcher to ensure a more robust scale and model development. The PCA of the pilot test was used to refine the research model by ensuring only adequate measures and groupings of factors were adopted for the overall model assessment. The output from this was a refined structural model that was subjected to reliability and validity testing.

**Reliability and Validity**

The presence of a high relative internal association amongst the research measured variables provides levels of internal reliability consistency (Trochim & Donnelly, 2006). Internal reliability is measured through the Cronbach's Alpha (CA) which is used to estimate the reliability of multi-scale items in research (Zikmund et al., 2009). However, Chin (2010) argues that within the SEM doctrine, the CA tends to underestimate the levels of reliability. Another measure, namely Composite Reliability (CR), should be adopted instead of CA, for internal reliability. CR takes the factor loadings of measured variables within the scale, into consideration (Chin, 2010). Both these indicators range from 0 to 1, whereby a minimum threshold of 0.7 is required to ascertain fair levels of reliability of a composite variable (Hair et al., 2010).

Convergent and discriminant validity of the hypothesized model was assessed using methods described by Hair et al. (2017). The extent in which an item can be associated, with other items within a component, is referred to as convergent validity, whereby these items need to have a common high measure of variance (Hulland, 1999). Hair et al. (2017), states that for convergent validity to be achieved, the measured variables factor loadings, on the respective latent variables, needs to exceed 0.708. In addition, the Average Variance Extracted (AVE), which is the square root of the standardised indicator loading, should be greater than 0.5 for convergent validity. An AVE less than 0.5 indicates that more errors remain on the item than the variance through the item's latent variable (Hair et al., 2017).

The Heterotrait-monotrait (HTMT) criterion measures the extent to which a hypothesised construct differs from other constructs in a research model (Chin, 2010). In essence, the HTMT criterion looks to provide an estimate of the real correlation that would exist between constructs if it were possible to be ideally measured (Henseler, Ringle, & Sarstedt, 2015). Thus, the research adopted the HTMT criterion to ensure no discriminant validity issues in the research model. If correlations between constructs are less than 0.9, this implies that the specific construct is relatively unique from the others and captures its hypothesized essence, which is not represented by other constructs in the research model (Hair et al., 2017; Henseler et al., 2015).

**Model assessment**

As the PLS technique is rooted in ordinary least square prediction and the maximization of dependent variable variance, the levels of collinearity need to be assessed (Hair et al., 2017). Collinearity is characterised by highly correlated independent variables, and high levels of collinearity can lead to predictor variable bias and create unstable path linkages (Chin, 2010). Henseler et al. (2015), recommends the assessment of the Variation Inflation Factor (VIF), which is interpreted as the inverse of tolerance. Furthermore, both (Henseler et al., 2015) and Hair et al. (2017), argue that the upper limit of VIF is either 10 or 5, respectively. Based on this differing academic view, the researcher adopted an upper limit of 5 to assess collinearity of the research model. If high levels of collinearity exist, those indicators will be deleted, and the model assessed for reliability and validity again.

Once the model was assessed for reliability, validity and collinearity, the path models were assessed for significance. The inner path coefficients between the research variables range from -1 to 1, whereby a negative value indicates a negative relationship and a positive value indicates a direct relationship. Additionally, the intensity of the path coefficient is a measure of strength between the hypothesised constructs, where -1 or 1 are the maximum limits and 0 the minimum. The significance of each path weighting was assessed using the bootstrap technique with a chosen significance level of 99%.

Model fit for the research model was verified by assessing the Standard Root Mean Square (SRMR), Coefficient of Determination ($R^2$), and Stone-Geisser's $Q^2$. SRMR is an evaluation between the discrepancies of the observed and expected relationships (Hair et al., 2017). Hair et al. (2017) recommend an upper limit of 0.10 for a good model fit. The $R^2$ measures the levels of variance between interaction variables and is a measure of the model's ability to predict, with a range from 0 to 1, whereby a value closer to 1 indicates high predictiveness (Roldán & Sánchez-Franco, 2012). According to (Henseler et al., 2015), a $R^2$ value larger than 0.67 is classified as substantial, between 0.19 and 0.33 is moderate, and less than 0.19 is weak.

The Stone-Geisser indicator evaluates the relative predictive relevance, especially suited for a reflective model (Roldán & Sánchez-Franco, 2012). The Stone-Geisser value ranges from 0 to 1 whereby values greater than 0.35 is classified as large significance, between 0.02 and 0.15 is medium, and less than 0.02 is classified as weak (Hair et al., 2017).

Based on the approach outlined, the researcher set out to develop a measurement scale for the determination of the practices the enable an EDLS. Given this context, the approach supported this goal through factor analysis. Both the inner and outer models were subjected to reliability and validity testing. From this, a final PLS model structure was confirmed and the significance of the path model assessed, based on the sample data available. This model was then subjected to an overall model fit assessment, confirming the accuracy in achieving the research goal.

### 2.9 Limitations

The methodology described was based on the aim of scale development and the output of a conceptual model given the requirement for achieving an EDLS. Based on this approach, the following limitations of the methodology were acknowledged as part of the research.

The cross-sectional time horizon exposed the research to varying degrees of maturity in organisations, that are intending to build data-led strategies. This potentially created a difference in respondent views given their exposure. Less mature organisations may potentially have not seen the benefits of BD as yet.

While the sampling methodology of the research lends itself to be purposive, the electronic medium used for survey distribution may result in a case of snowballing and create bias in the received results. Control questions were used to minimise the impact however, this could not entirely be mitigated. It relied on the honesty of the respondents, based on their experience with both BD and strategy.

The sample obtained was comprised of respondents across multiple countries, however predominantly South African based. Given this limitation, there could be skewness present in the results due to the varying maturity associated with BDAC within the South African landscape. Furthermore, the geographic distribution of the population was not diverse or large enough to provide a generalised view on EDLS practices for managers.

In the context of the analysis and scale development, the PLS-SEM technique provides a more robust statistical method versus the first order statistical techniques but only reports on a single fit index (SRMR). While this is acceptable, the maturity of PLS-SEM needs to be improved to provide more fit indices similar to more developed methods (Hair et al., 2017).

The constructs at a second order and first order level were modelled as independent antecedents towards an EDLS respectively. While this is based on theoretical underpinnings, there needs to be further investigation into the relationships that exist between each of the constructs and how this may influence the practice of EDLS. Potentially leveraging the theoretical underpinnings of modes of adaptation, the theory of behavioural integration (BI) considers management of conflicting or separate activities in a single business entity (Birkinshaw, Zimmermann, & Riasch, 2016). In addition to this Birkinshaw et al. (2016), also propose the theories of structural separation and sequential alternation as mechanisms to manage the relationships of factors towards achieving a practice for EDLS.

Finally, the research by Akter et al. (2016) proposed a metric to measure the impact of BDAC through firm performance. This provided an ability to assess the value created as a result of BDAC, through the levers of market performance and financial performance or operational performance (Gupta & George, 2016; Wamba et al., 2017). Further investigation should be conducted to determine the extent of value creation through EDLS with relevant metrics such as firm performance suggested above, to assess the effectiveness of the proposed factors from a business perspective.

## 3. Reference List


Abbasi, A., Sarker, S., & Chiang, R. H. L. (2016). Big data research in information systems: Toward an inclusive research agenda. *Journal of the Association of Information Systems*, *17*(2), 1–32.

Adrian, C., Abdullah, R., Atan, R., & Jusoh, Y. Y. (2018). Conceptual Model Development of Big Data Analytics Implementation Assessment Effect on Decision-Making. *International Journal of Interactive Multimedia and Artificial Intelligence*, *5*(1), 101–106. https://doi.org/10.9781/ijimai.2018.03.001

Akter, S., Wamba, S. F., Gunasekaran, A., Dubey, R., & Childe, S. J. (2016). How to improve firm performance using big data analytics capability and business strategy alignment? *International Journal of Production Economics*, *182*, 113–131. https://doi.org/10.1016/j.ijpe.2016.08.018

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, *17*(1), 99–120. https://doi.org/0803973233

Birkinshaw, J., Zimmermann, A., & Riasch, S. (2016). How Do Firms Adapt to Discontinuous Change? *California Management Review*, *58*(4), 36–58. https://doi.org/10.1525/cmr.2016.58.4.36

Braeken, J., & Van Assen, M. A. L. M. (2017). An empirical Kaiser criterion. *Psychological Methods*, *22*(3), 450–466. https://doi.org/10.1037/met0000074

Brattström, A., Frishammar, J., Richtnér, A., & Pflueger, D. (2018). Can innovation be measured? A framework of how measurement of innovation engages attention in firms. *Journal of Engineering and Technology Management - JET-M*, *48*(March), 64–75. https://doi.org/10.1016/j.jengtecman.2018.04.003

Camisón, C., & Forés, B. (2010). Knowledge absorptive capacity: New insights for its conceptualization and measurement. *Journal of Business Research*, *63*(7), 707–715. https://doi.org/10.1016/j.jbusres.2009.04.022

Carpenter, S. (2018). Ten Steps in Scale Development and Reporting: A Guide for Researchers. *Communication Methods and Measures*, *12*(1), 25–44. https://doi.org/10.1080/19312458.2017.1396583

Carruthers, C., & Jackson, P. (2018). *The Chief Data Officer's Playbook*. London: Facet Publishing.

Casadesus-Masanell, R., & Ricart, J. (2009). *From Strategy to Business Models and to Tactics* (036 No. 10). Boston, Massachusetts. https://doi.org/10.1016/j.lrp.2010.01.004

Chang, W., Franke, G. R., & Lee, N. (2016). Comparing reflective and formative measures: New insights from relevant simulations. *Journal of Business Research*, *69*(8), 3177–3185. https://doi.org/10.1016/j.jbusres.2015.12.006

Chen, P. C. L., & Zhang, C.-Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*, *275*, 314–347. https://doi.org/10.1016/j.ins.2014.01.015

Chen, Y. S., Lin, M. J. J., & Chang, C. H. (2009). The positive effects of relationship learning and absorptive capacity on innovation performance and competitive advantage in industrial markets. *Industrial Marketing Management*, *38*(2), 152–158. https://doi.org/10.1016/j.indmarman.2008.12.003

Chin, W. W. (2010). How to write up and report PLS analyses. In *In Handbook of partial least squares* (pp. 655–690). Springer Handbooks of Computational Statistics. Retrieved from https://doi.org/10.1007/978-3-540-32827-8_29

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly*, *35*(1), 128–152. https://doi.org/10.2307/2393553

Costello, A. B., & Osborne, J. W. (2005). Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis Anna. *Practical Assessment, Research & Evaluation*, *10*(7), 1–9. Retrieved from papers2://publication/uuid/256DEBAD-ECC4-4DC4-99FD-E67E4D567AEC

Creswell, J. W. (2014). *Research Design: Qualitative, Quantitative and Mixed Methods Approaches* (Fourth Edi). Washington DC: Sage Publications, Inc.

de Wet, P. (2019). The definitions of micro, small, and medium businesses have just been radically overhauled – here's how. Retrieved June 15, 2019, from https://www.businessinsider.co.za/micro-small-and-medium-business-definition-update-by-sector-2019-3

Dubey, R., Gunasekaran, A., Childe, S. J., Blome, C., & Papadopoulos, T. (2019). Big Data and Predictive Analytics and Manufacturing Performance: Integrating Institutional Theory, Resource-Based View and Big Data Culture. *British Journal of Management*, *30*(2), 341–361. https://doi.org/10.1111/1467-8551.12355

Dubey, R., Gunasekaran, A., Childe, S. J., Roubaud, D., Wamba, S. F., Giannakis, M., & Foropon, C. (2019). Big data analytics and organizational culture as complements to swift trust and collaborative performance in the humanitarian supply chain. *International Journal of Production Economics*, *210*(June 2018), 120–136. https://doi.org/10.1016/j.ijpe.2019.01.023

Ferreira, L. C. de M. (2017). Sense and sensibility: Testing an attention-based view of organizational responses to social issues. *Business Ethics*, *26*(4), 443–456. https://doi.org/10.1111/beer.12156

Fink, T. M. A., Reeves, M., Palma, R., & Farr, R. S. (2017). Serendipity and strategy in rapid innovation. *Nature Communications*, *8*(1), 1–9. https://doi.org/10.1038/s41467-017-02042-w

Frisk, J. E., & Bannister, F. (2017). Improving the use of analytics and big data by changing the decision-making culture: A design approach. *Management Decision*, *55*(10), 2074–2088. https://doi.org/10.1108/MD-07-2016-0460

Gebauer, H. (2009). An attention-based view on service orientation in the business strategy of manufacturing companies. *Journal of Managerial Psychology*, *24*(1), 79–98. https://doi.org/10.1108/02683940910922555

Gemma, R. (2018). Introduction to positivism, interpretivism and critical theory. *Nurse Researcher*, *25*(4), 14–20. https://doi.org/10.7748/nr.2018.e1466

Ghasemaghaei, M. (2019). Are firms ready to use big data analytics to create value? The role of structural and psychological readiness. *Enterprise Information Systems*. https://doi.org/10.1080/17517575.2019.1576228

Ghasemaghaei, M., Hassanein, K., & Turel, O. (2017). Increasing firm agility through the use of data analytics: The role of fit. *Decision Support Systems*, *101*, 95–105. https://doi.org/10.1016/j.dss.2017.06.004

Grant, R. M. (1991). The Resource-Based Theory of Competitive Advantage: ImpHcations for Strategy Formulation Figure 1. A Resource-Based Approach to Strategy Analysis: A Practical Framework. *California Management Review*, *33*(3), 114–135. https://doi.org/10.2307/41166664.

Gupta, M., & George, J. F. (2016). Toward the development of a big data analytics capability. *Information and Management*, *53*(8), 1049–1064. https://doi.org/10.1016/j.im.2016.07.004

Hair, J. A., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2016). *A primer on partial least squares structural equation modelling (PLS-SEM)* (2nd ed.). Sage.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis: A Global Perspective* (Seventh Ed). New Jersey: Pearson Education.

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS SEM Indeed a Silver Bullet.pdf. *Journal of Marketing Theory and Practice*, *19*(2), 139–151. https://doi.org/10.2753/ MTP1069-6679190202

Hair, J. F., Sarstedt, M., Hopkins, L., & Kuppelwieser, V. G. (2014). Partial least squares structural equation modeling (PLS-SEM): An emerging tool in business research. *European Business Review*, *26*(2), 106–121. https://doi.org/10.1108/EBR-10-2013-0128

Hair Jr., J. F., Matthews, L. M., Matthews, R. L., & Sarstedt, M. (2017). PLS-SEM or CB-SEM: updated guidelines on which method to use. *International Journal of Multivariate Data Analysis*, *1*(2), 107–123. https://doi.org/10.1504/ijmda.2017.10008574

Hardin, A. M., Chang, J. C.-J., & Fuller, M. A. (2008). Formative vs. Reflective Measurement: Comment on Marakas, Johnson, and Clay (2007). *Journal of the Association for Information Systems*, *9*(9), 535–543.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Hoyle, R. H. (1995). The structural equation modeling approach: Basic concepts and fundamental issues. In *Structural equation modeling: Concepts, issues, and applications.* (pp. 1–15). Thousand Oaks, CA, US: Sage Publications, Inc.

Hulland, J. (1999). Use of partial least squares (PLS) in strategic management research: a review of four recent studies. *Strategic Management Journal*, *20*(2), 195–204. https://doi.org/10.1002/(sici)1097-0266(199902)20:2<195::aid-smj13>3.3.co;2-z

Johanson, G. A., & Brooks, G. P. (2010). Initial scale development: Sample size for pilot studies. *Educational and Psychological Measurement*, *70*(3), 394–400. https://doi.org/10.1177/0013164409355692

Joseph, J., & Ocasio, W. (2012). Architecture, Attention, and Adaptation in the Multibusiness Firm: General Electric from 1951 to 2001. *Strategic Management Journal*, *33*(1), 633–660. https://doi.org/10.1002/smj

Joseph, J., & Wilson, A. J. (2018). The growth of the firm: An attention-based view. *Strategic Management Journal*, *39*, 1779–1800. https://doi.org/10.1002/smj.2715

Kitchens, B., Dobolyi, D., Li, J., & Abbasi, A. (2018). Advanced Customer Analytics: Strategic Value Through Integration of Relationship-Oriented Big Data. *Journal of Management Information Systems*, *35*(2), 540–574. https://doi.org/10.1080/07421222.2018.1451957

Koryak, O., Lockett, A., Hayton, J., Nicolaou, N., & Mole, K. (2018). Disentangling the antecedents of ambidexterity: Exploration and exploitation. *Research Policy*, *47*(2), 413–427. https://doi.org/10.1016/j.respol.2017.12.003

Lee, I. (2017). Big data : Dimensions , evolution , impacts , and challenges. *Business Horizons*, *60*(3), 293–303. https://doi.org/10.1016/j.bushor.2017.01.004

Martin, R. (2014). Adaptive Strategy Is a Cop-Out. Retrieved April 12, 2019, from http://web.b.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=83f654a9-a93e-4ae1-b8ad-cc90af9c452d@sessionmgr103&vid=1&hid=130

Mazzei, M. J., & Noble, D. (2017). Big data dreams: A framework for corporate strategy. *Business Horizons*, *60*(3), 405–414. https://doi.org/10.1016/j.bushor.2017.01.010

Mcabee, S. T., Landis, R. S., & Burke, M. I. (2017). Inductive reasoning : The promise of big data. *Human Resource Management Review*, *27*(2), 277–290. https://doi.org/10.1016/j.hrmr.2016.08.005

McAfee, A., & Brynjolfsson, E. (2012). Big Data: The Management Revolution. https://doi.org/10.3823/5004

Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2019). Big data analytics capabilities and innovation: The mediating role of dynamic capabilities and moderating effect of the environment. *British Journal of Management*, *30*, 272–298. https://doi.org/10.1111/1467-8551.12343

Mintzberg, H. (1987). The Stratergy Concept I: Five Ps For. *California Management Review*, *30*(1), 11–24. Retrieved from https://slack-files.com/files-pri-safe/TDETUV0LW-FDEG2RNHJ/mintzberg_1987_strategyasconcept_prereading1.pdf?c=1540116897-f82939966bb3525c5c059d618d6ef37f58ac40d4

Newman, D. A. (2014). Missing Data : Five Practical Guidelines. *Organizational Research Methods*, *17(4)*(372–411). https://doi.org/10.1177/1094428114548590

Ocasio, W. (1997). Towards an Attention Based View of the Firm. *Strategic Management Journal*, *18*, 187–206.

Ocasio, W., & Joseph, J. (2005). An Attention-Based Theory of Strategy Formulation: Linking Micro- and Macroperspectives in Strategy Processes. *Advances in Strategic Management*, *22*, 39–61. https://doi.org/10.1016/S0742-3322(05)22002-8

Ocasio, W., & Joseph, J. (2018). The Attention-Based View of Great Strategies . *Strategy Science*, *3*(1), 289–294. https://doi.org/10.1287/stsc.2017.0042

Ocasio, W., Laamanen, T., & Vaara, E. (2018). Communication and attention dynamics: An attention-based view of strategic change. *Strategic Management Journal*, *39*, 155–167. https://doi.org/10.1002/smj.2702

Pallant, J. (2007). *SPSS survival manual* (Third). New York: McGraw Hill.

Palmié, M., Lingens, B., & Gassmann, O. (2016). Towards an attention-based view of technology decisions. *R&D Management*, *46*(4), 781–796.

Parnell, J. A., & Lester, D. L. (2003). Towards a philosophy of strategy: reassessing five critical dilemmas in strategy formulation and change. *Strategic Change*, *12*(6), 291–303. https://doi.org/10.1002/jsc.639

Raguseo, E., & Vitari, C. (2018). Investments in big data analytics and firm performance: an empirical investigation of direct and mediating effects. *International Journal of Production Research*, *56*(15), 5206–5221. https://doi.org/10.1080/00207543.2018.1427900

Raisch, S., Birkinshaw, J., Probst, G., & Tushman, M. L. (2009). Organizational Ambidexterity: Balancing Exploitation and Exploration for Sustained Performance. *Organization Science*, *20*(4), 685–695. https://doi.org/10.1287/orsc.1090.0428

Rigdon, E. E., Sarstedt, M., & Ringle, C. M. (2017). On Comparing Results from CB-SEM and PLS-SEM: Five Perspectives and Five Recommendations. *Journal of Research and Management*, *39*(3), 4–16. https://doi.org/10.15358/0344-1369-2017-3-4

Rivard, S., Raymond, L., & Verreault, D. (2006). Resource-based view and competitive strategy: An integrated model of the contribution of information technology to firm performance. *Journal of Strategic Information Systems*, *15*(1), 29–50. https://doi.org/10.1016/j.jsis.2005.06.003

Roberts, N., Galluch, P. S., Dinger, M., & Grover, V. (2012). Absorptive Capacity and Information Systems Research: Review, Synthesis, and Directions for Future Research. *MIS Quarterly*, *36*(2), 625–648. Retrieved from http://www.misq.org

Rodriguez, L., & Da Cunha, C. (2018). Impacts of Big Data Analytics and Absorptive Capacity on Sustainable Supply Chain Innovation: A Conceptual Framework. *Scientific Journal of Logisitics*, *14*(2), 151–161. https://doi.org/http://dx.doi.org/10.17270/J.LOG.267

Roldán, J. L., & Sánchez-Franco, M. J. (2012). *Variance-based structural equation modeling: Guidelines for using partial least squares in information* *systems research*. *Research Methodologies, Innovations and Philosophies in Software Systems Engineering and Information Systems*. https://doi.org/10.4018/978-1-4666-0179-6.ch010

Rude, D. A. (2014). New Leader Agendas & Legacy Organizational Culture: A Case Study Exploring Learning's Role in Bridging the Great Divide. *American Journal of Management*, *14*(1–2), 82–101. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=100405143&site=ehost-live

Saunders, M., & Lewis, P. (2012). *Doing Research in Business and Managment* (Second). Pearson.

Scheffer, J. (2002). Dealing with missing data in family-based association studies: A multiple imputation approach. *Research Letters in the Information and Mathematical Sciences*, *3*, 153–160. https://doi.org/10.1159/000100481

Schein, E. H. (2004). *Organizational Culture and Leadership* (Third). San Francisco: Jossey-Bass.

Solís-Molina, M., Hernández-Espallardo, M., & Rodríguez-Orejuela, A. (2018). Performance implications of organizational ambidexterity versus specialization in exploitation or exploration: The role of absorptive capacity. *Journal of Business Research*, *91*(September 2017), 181–194. https://doi.org/10.1016/j.jbusres.2018.06.001

Soo, C., Wei Tian, A., Teo, S. T. T., & Cordery, J. (2017). Intellectual Capital-Enhancing HR,Absorptive Capacity, and Innovation. *Human Resource Management*, *56*(3), 431–454. https://doi.org/10.1002/hrm

Srivastava, U., & Gopalkrishnan, S. (2015). Impact of big data analytics on banking sector: Learning for Indian Banks. *Procedia Computer Science*, *50*, 643–652. https://doi.org/10.1016/j.procs.2015.04.098

Tawse, A., Patrick, V. M., & Vera, D. (2019). Crossing the chasm: Leadership nudges to help transition from strategy formulation to strategy implementation. *Business Horizons*, *62*(2), 249–257. https://doi.org/10.1016/j.bushor.2018.09.005

Trochim, W. M., & Donnelly, J. P. (2006). *The Research Methods Knowledge*

*Base* (Third). Cincinnati, OH: Cengage Learning.

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J. fan, Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, *70*, 356–365. https://doi.org/10.1016/j.jbusres.2016.08.009

Wang, H., Xu, Z., Fujita, H., & Liu, S. (2016). Towards felicitous decision making: An overview on challenges and trends of Big Data. *Information Sciences*, *367–368*, 747–765. https://doi.org/10.1016/j.ins.2016.07.007

Wixom, B. H., Yen, B., & Relich, M. (2013). Maximizing Value from Business Analytics. *MIS Quarterley Executive*, *12*(2), 111–123.

Xie, X., Zou, H., & Qi, G. (2018). Knowledge absorptive capacity and innovation performance in high-tech companies: A multi-mediating analysis. *Journal of Business Research*, *88*(June 2017), 289–297. https://doi.org/10.1016/j.jbusres.2018.01.019

Yli-Renko, H., Autio, E., & Sapienza, H. J. (2001). Social capital, knowledge acquisition, and knowledge exploitation in young technology-based firms. *Strategic Management Journal*, *22*(6–7), 587–613. https://doi.org/10.1002/smj.183

Zahra, S. A., & George, G. (2002). Absorptive Capacity : A Review , Reconceptualization , and Extension . *Academy of Management Review*, *27*(2), 185–203. https://doi.org/https://doi.org/10.5465/AMR.2002.6587995

Zikmund, W. G., Babin, B. J., Carr, J. C., & Griffin, M. (2009). *Business Research Methods*. *Business Research Methods* (Eight). Cengage Learning. https://doi.org/10.1016/S1872-5813(14)60037-2

## 4. Appendix A

Manuscripts submitted to *Business Horizons* should address topical and timely issues of relevance to business academicians and practitioners. Successful submissions will typically be structured around identifying and developing a problem or issue and providing relevant solutions. Importantly, manuscripts should go beyond description and offer sound prescriptive advice. Manuscripts should also be solidly grounded in a scholarly foundation with appropriate and judicious use of source citations. Manuscripts should also be written in clear, non-technical language, with a broad business readership in mind. While the language should be engaging and informative, authors should avoid the use of jargon and technical terminology.

Manuscripts should be prepared consistent with the following guidelines. Manuscripts which do not conform to these guidelines may be returned to the author(s) without review for reformatting.

1. Double-space, use a 12 point font with normal text spacing, and one-inch margins throughout the entire manuscript. Manuscripts should not exceed 25 pages, all-inclusive. We cannot, however, consider notes, briefs, or commentaries. All pages, save the title page, should include pagination. Page numbers should appear centered at the bottom of each page.

2. The first page of the manuscript should include the title of the manuscript and complete contact information for each author with author name, affiliation, full postal mail address, email address, telephone number, and fax number. The corresponding author should be clearly noted in the case of multiple authors.

3. The second page of the manuscript should include the title of the manuscript, an abstract of 150 to 200 words, and four to five key words or short phrases that accurately reflect the content of the manuscript. Abstracts should be designed to provide a comprehensive executive summary of the manuscript in a manner that draws the reader's attention.

4. The body of the text should begin on the third manuscript page. The manuscript text should begin with an introductory heading.

5. Incorporate headings and sub-headings throughout the manuscript to aid readability. First order headings should be centered and all capital letters. Second order headings should be centered and use both upper and lower case letters. Headings should be descriptive and informative, yet not standard academic style.

For example, rather than use "Introduction", you might elect to use "Corporate Women: Another Look". The aim is to guide the reader with innovative and lively language.

6. *Business Horizons* relies on the APA (American Psychological Association) style of referencing. Authors should carefully document their work while at the same time judiciously select references. A complete list of references cited should appear at the end of the text, and preceding any tables, figures, or graphs. Only works cited in the manuscript should be included in the references section. The references should begin on a new manuscript page, with the heading REFERENCES appearing centered at the top of this page. We do not rely on footnotes or endnotes. Any full or partial with-in text quoted material should include the relevant page number(s) with the source citation (e.g., Author & Author, 2008, p.1). Please also note the use of ampersand for within text citations contained in parentheses. The following are examples of the APA referencing style. Please consult the *Publication Manual of the American Psychological Association,* Fifth Edition, ISBN 1-55798-790-4 or [http://www.apastyle.org/elecref.html](http://www.apastyle.org/elecref.html) for further style guidelines.

Data references

This journal encourages you to cite underlying or relevant datasets in your manuscript by citing them in your text and including a data reference in your Reference List. Data references should include the following elements: author name(s), dataset title, data repository, version (where available), year, and global persistent identifier. Add [dataset] immediately before the reference so we can properly identify it as a data reference. This identifier will not appear in your published article.

**Journal article:**

Stuart, F. I. (2006). Designing and executing memorable service experiences: Lights, camera, *experiment, integrate*, action! *Business Horizons, 49*(2), 149-159.

Ketchen, D., & Hult, G. T. (2007). Bridging organization theory and supply chain management: The case of best value supply chains. *Journal of Operations Management, 25*(2), 573-580.

**Book:**

Miller, D., & Le Breton-Miller, I. (2005). *Managing for the long run: Lessons in* *competitive advantage from great family businesses.* Boston: Harvard Business School Press.

Cialdini, R. B. (2001). *Influence: Science and practice* (4th ed.). Boston: Allyn & Bacon.

**Edited collection:**

Pfeffer, J. (1998). Understanding organizations: Concepts and controversies. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (pp. 733-777). New York: McGraw-Hill.

Geis, G. (1982). The heavy electrical equipment anti-trust cases of 1961. In M. D. Ermann & R. J. Lundman (Eds.), *Corporate and governmental deviance* (pp. 123-143). New York: Oxford University Press.

**Web source:**

Berry, L. L., & Seltman, K. D. (2007). Building a strong services brand: Lessons from Mayo Clinic. *Business Horizons, 50*(3), 199-209. Retrieved May 10, 2007, from https://www.sciencedirect.com.

Lewis, P. H. (2007). *How Apple kept its iPhone secrets.* Retrieved January 31, 2008, from http://money.cnn.com/2007/01/10/commentary/lewis_fortune_iphone.fortune/index.htm

**Dataset:**

Oguro, M., Imahiro, S., Saito, S., Nakashizuka, T. (2015). *Mortality data for Japanese oak wilt disease and surrounding forest compositions.* Mendeley Data, v1. https://doi.org/10.17632/xwj98nb39r.1.

7. All tables, figures, graphs, or appendices should appear individually on a separate manuscript page. These should clarify or supplement the manuscript text, not duplicate the text. Please indicate the appropriate placement of tables, figures, and graphs within the text by using [Insert Table 1 about here] place on a separate text line. These should appear following the References in the following order: tables, figures, graphs, appendices.

8. Authors should carefully proofread their manuscripts prior to submission. Please pay careful attention to spelling and grammar, in particular. Also, please rely on gender neutral language. Manuscripts with extensive errors will be returned without review. Submission of a manuscript to *Business Horizons* implies a commitment by the author(s) to engage in the review process and to have the article published

should it be accepted. Articles previously published, those under consideration by another journal, and those with a pre-existing copyright may not be submitted for review and consideration. Upon submission, authors also agree not to submit the manuscript for consideration elsewhere during the review period.

This journal operates a double blind review process. All contributions will be initially assessed by the editor for suitability for the journal. Papers deemed suitable are then sent to independent expert reviewer(s) to assess the scientific quality of the paper. The Editor is responsible for the final decision regarding acceptance or rejection of articles. The Editor's decision is final. More information on types of peer review.

9. **Submit all manuscripts electronically in MS Word-compatible file format to the editor at bushor@indiana.edu.** Submissions will be acknowledged shortly after receipt. Editorial decisions may take up to three months. All editorial decisions are final. Articles typically appear in print 8 to 12 months after final acceptance.

10. Materials published in *Business Horizons* are available for viewing and download via ScienceDirect.

11. In order to afford all authors an opportunity of publication, authors are limited to a maximum of three submissions per calendar year.

12. This journal offers authors a choice in publishing their research: Open Access and Subscription.

*Source: Authors guidelines. Sourced from "Guide for Authors: Business Horizons," by Elsevier, 2019, Retrieved from https://www.elsevier.com/journals/business-horizons/0007-6813/guide-for-authors*

**Gavin Reuben** 19 August 2019 at 19:51

GR

Re: [External] Enquiry: PLS SEM Articles

To: Greg Fisher

Hi Greg,

To ensure I fully understand the manuscript author guidelines:

What font style is recommended? It mentions a font size and line spacing but no guideline on the text?

It mentions a maximum of 25 pages - Does this include references, title page and author details as well as the questionnaire as an appendix, in a quant study?
Finally, should the manuscript be submitted as a single column text document? One column of text per a page?

Thanks for your guidance and support in this regard,
Gavin

**See More** from Greg Fisher

**Greg Fisher** 19 August 2019 at 20:08

GF

Re: [External] Enquiry: PLS SEM Articles

To: Gavin Reuben

Gavin

There is no specified type of font. I recommend Times New Roman. 25 pages is the text. Other elements don't count toward that. A single column is correct.

Greg

**See More** from Gavin Reuben

--
**Greg Fisher**
Larry and Barbara Sharpf Professor of Entrepreneurship
Associate Professor
Kelley School of Business
Indiana University

Editor: Business Horizons
Field Editor: Journal of Business Venturing

Phone: (812) 855-2763
Email: fisherg@indiana.edu

*Figure 3: Journal guideline confirmation*

## 5. Appendix B

Business Horizons (2017) 60, 293—303

Available online at www.sciencedirect.com

ScienceDirect

KELLEY SCHOOL OF BUSINESS

INDIANA UNIVERSITY

www.elsevier.com/locate/bushor

# Big data: Dimensions, evolution, impacts, and challenges

In Lee

*School of Computer Sciences, Western Illinois University, 1 University Circle, Macomb, IL 61455-1390, U.S.A.*

**KEYWORDS**
Big data;
Internet of things;
Data analytics;
Sentiment analysis;
Social network analysis;
Web analytics

**Abstract** Big data represents a new technology paradigm for data that are generated at high velocity and high volume, and with high variety. Big data is envisioned as a game changer capable of revolutionizing the way businesses operate in many industries. This article introduces an integrated view of big data, traces the evolution of big data over the past 20 years, and discusses data analytics essential for processing various structured and unstructured data. This article illustrates the application of data analytics using merchant review data. The impacts of big data on key business performances are then evaluated. Finally, six technical and managerial challenges are discussed.


## 1. The day of big data

The emerging technological development of big data is recognized as one of the most important areas of future information technology and is evolving at a rapid speed, driven in part by social media and the Internet of Things (IoT) phenomenon. The technological developments in big data infrastructure, analytics, and services allow firms to transform themselves into data-driven organizations. Due to the potential of big data becoming a game changer, every firm needs to build capabilities to leverage big data in order to stay competitive. IDC (2015) forecasted that the big data technology and services market will grow at a compound annual growth rate of 23.1% over the 2014—2019 period, with annual spending reaching $48.6 billion in 2019.

While structured data is an essential part of big data, more and more data are created in unstructured video and image forms, which traditional data management technologies cannot process adequately. A large portion of data worldwide have been generated by billions of IoT devices such as smart home appliances, wearable devices, and environmental sensors. Gartner (2015) forecasted that 4.9 billion connected objects would be in use in 2015—up 30% from 2014—and will reach 25 billion by 2020. To meet the ever-increasing storage and processing needs of big data, new big data platforms

*E-mail address:* i-lee@wiu.edu

0007-6813/$ — see front matter 
http://dx.doi.org/10.1016/j.bushor.2017.01.004

are emerging, including NoSQL[1] databases as an alternative to traditional relational databases and Hadoop as an open source framework for inexpensive, distributed clusters of commodity hardware.

In this article, I start with a discussion of big data dimensions and trace the evolution of big data since 1995. Then, I illustrate the application of data analytics using a scenario involving merchant review data. In the following section, I discuss impacts of big data on various business performances. Finally, I discuss six technical and managerial challenges: data quality, data security, privacy, data management, investment justification, and shortage of qualified data scientists.

## 2. Dimensions of big data

Laney (2001) suggested that volume, variety, and velocity are the three dimensions of big data. The 3 Vs have been used as a common framework to describe big data (Chen, Chiang, & Storey, 2012; Kwon, Lee, & Shin, 2014). Here, I describe the 3 Vs and additional dimensions of big data proposed in the computing industry.

*Volume* refers to the amount of data an organization or an individual collects and/or generates. While currently a minimum of 1 terabyte is the threshold of big data, the minimum size to qualify as big data is a function of technology development. Currently, 1 terabyte stores as much data as would fit on 1,500 CDs or 220 DVDs, enough to store around 16 million Facebook photographs (Gandomi & Haider, 2015). E-commerce, social media, and sensors generate high volumes of unstructured data such as audio, images, and video. New data has been added at an increasing rate as more computing devices are connected to the internet.

*Velocity* refers to the speed at which data are generated and processed. The velocity of data increases over time. Initially, companies analyzed data using batch processing systems because of the slow and expensive nature of data processing. As the speed of data generation and processing increased, real time processing became a norm for computing applications. Gartner (2015) forecasted that 6.4 billion connected devices would be in use worldwide in 2016 and that the number will reach 20.8 billion by 2020. In 2016, 5.5 million new devices were estimated to be connected every day to collect, analyze, and share data. The enhanced data streaming capability of connected devices will continue to accelerate the velocity.

[1] Interpreted as Not Only SQL

*Variety* refers to the number of data types. Technological advances allow organizations to generate various types of structured, semi-structured, and unstructured data. Text, photo, audio, video, clickstream data, and sensor data are examples of unstructured data, which lack the standardized structure required for efficient computing. Semi-structured data do not conform to specifications of the relational database, but can be specified to meet certain structural needs of applications. An example of semi-structured data is Extensible Business Reporting Language (XBRL), developed to exchange financial data between organizations and government agencies. Structured data is predefined and can be found in many types of traditional databases. As new analytics techniques are developed, unstructured data are generated at a much faster rate than structured data and the data type becomes less of an impediment for the analysis.

IBM added *veracity* as a fourth dimension, which represents the unreliability and uncertainty latent in data sources. Uncertainty and unreliability arise due to incompleteness, inaccuracy, latency, inconsistency, subjectivity, and deception in data. Managers do not trust data when veracity issues are prevalent. Customer sentiments are unreliable and uncertain due to subjectivity of human opinions. Statistical tools and techniques have been developed to deal with uncertainty and unreliability of big data with specified confidence levels or intervals.

SAS added two additional dimensions to big data: variability and complexity. *Variability* refers to the variation in data flow rates. In addition to the increasing velocity and variety of data, data flows can fluctuate with unpredictable peaks and troughs. Unpredictable event-triggered peak data are challenging to manage with limited computing resources. On the other hand, investment in resources to meet the peak-level computing demand will be costly due to overall underutilization of the resources. *Complexity* refers to the number of data sources. Big data are collected from numerous data sources. Complexity makes it difficult to collect, cleanse, store, and process heterogeneous data. It is necessary to reduce the complexity with open sources, standard platforms, and real-time processing of streaming data.

Oracle introduced *value* as an additional dimension of big data. Firms need to understand the importance of using big data to increase revenue, decrease operational costs, and serve customers better; at the same time, they must consider the investment cost of a big data project. Data would be low value in their original form, but data analytics will transform the data into a high-value strategic asset. IT professionals need to assess the benefits and costs of collecting and/or generating big data, choose high-value data sources, and build analytics capable of providing value-added information to managers.

As discussed above, a number of dimensions were presented in the computing industry. These dimensions together help us understand big data. I propose one additional dimension to big data: decay. *Decay* of data refers to the declining value of data over time. In a time of high velocity, the timely processing and acting on analysis is all the more important. IoT devices generate high volumes of streaming data, and immediate processing is often required for time-critical situations such as patient monitoring and environmental safety monitoring. Wearable medical devices such as glucose monitors, pulse oximeters, and blood pressure monitors worn on or close to the body produce a stream of data on patients' physiological conditions. In the era of big data, the decay of data will be an exponential function of time.

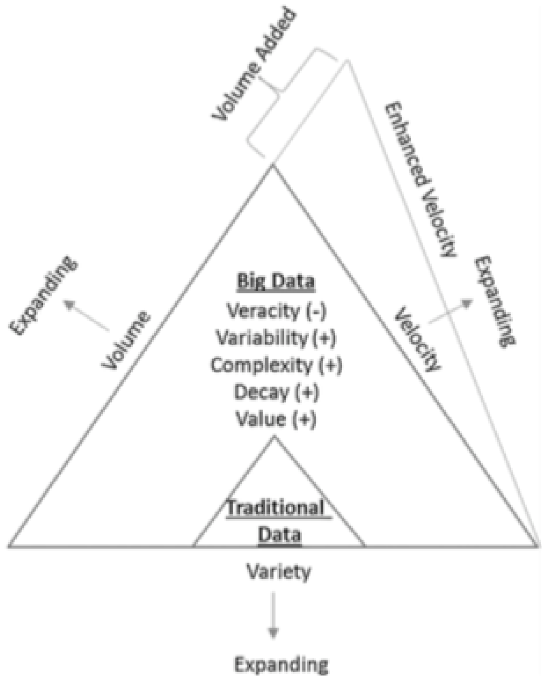
## 3. An integrated view of big data

Currently, the dimensions of big data have been proposed separately in the computing industry, but so far we lack an integrated view of big data. To help managers fully understand the relationships between the dimensions of big data, I present an integrated view of big data in Figure 1. The integrated view shows how these dimensions are interrelated with each other.

The three edges of the integrated view of big data represent three dimensions of big data: volume, velocity, and variety. Inside the triangle are the five dimensions of big data that are affected by the growth of the three triangular dimensions: veracity, variability, complexity, decay, and value. The growth of the three-edged dimensions is negatively related to veracity, but positively related to complexity, variability, decay, and value.

The integrated view shows that traditional data is a subset of big data with the same three dimensions, but the scope of each dimension is much smaller than that of big data. Traditional data consist mostly of structured data, and relational database management systems have been widely used to collect, store, and process the traditional data. As the scope of the three dimensions continues to expand, the proportion of the unstructured data increases. The magnitude of big data is expanding with the growth of velocity, volume, and variety. The arrows represent the expansion of each of the three dimensions. The expansion of velocity,

**Figure 1. An integrated view of big data**



volume, and variety are intertwined with each other. The expansion of each dimension affects the other seven dimensions. For example, Figure 1 shows that the expansion of velocity affects either volume or variety or both, and consequently affects the other five dimensions of big data inside the triangle (i.e., veracity declines, but variability, complexity, decay, and value increase).

## 4. Evolution of big data and data analytics

While the emergence of big data occurred only recently, the act of gathering and storing large amounts of data dates back to the early 1950s when the first commercial mainframe computers were introduced. During the period between the early 1950s to mid-1990s, data grew relatively slowly due to the high cost of computers, storage, and data networks. Data during this period were highly structured, mainly to support operational and transactional information systems. The advent of the world wide web (WWW) in the early 1990s led to the explosive growth of data and the development of big data analytics. Since the advent of the WWW, big data and data analytics have evolved through three major stages.

### 4.1. Big Data 1.0 (1994–2004)

Big Data 1.0 coincides with the advent of e-commerce in 1994, during which time online firms were the main contributors of the web content. User-generated content was only a marginal part of web content due to the technical limitation of web applications. In this era, web mining techniques were developed to analyze users' online activities. Web mining can be divided into three different types: web usage mining, web structure mining, and web content mining.

*Web usage mining* is the application of data mining techniques to discover web users' usage patterns online. Usage data captures the identity or origin of web users along with their browsing behavior. The ability to track individual users' mouse clicks, searches, and browsing patterns makes it possible to provide personalized services to users.

*Web structure mining* is the process of analyzing the structure of a website or a web page. The structure of a typical website consists of web pages as nodes and hyperlinks as edges connecting related pages. A hyperlink connects a location in a web page to a different location, either within the same web page or on a different web page. Based on the hyperlink structure, web pages are categorized. Google's PageRank, rooted in social network analysis, analyzes the hyperlink structure of web pages to rank them according to their degree of popularity or importance.

*Web content mining* is the process of extracting useful information from the content of web pages. A web page may consist of text, images, audio, video, or Extensible Markup Language-based (XML-based) data. Text mining has been applied widely to web content mining. Text mining extracts information from unstructured text and draws heavily on techniques from such disciplines as information retrieval (IR) and natural language processing (NLP). In its simplest form, text mining extracts a certain set of words or terms that are commonly used in the text. Web content mining is concerned about extraction of web page information, clustering of web pages, and classification of web pages into cyberterrorism, email fraud, spam mail filtering, etc. While there existed mining techniques in image processing and computer vision, the application of these techniques to web content mining was limited during the Big Data 1.0 era.

### 4.2. Big Data 2.0 (2005–2014)

Big Data 2.0 is driven by Web 2.0 and the social media phenomenon. Web 2.0 refers to a web paradigm that evolved from the web technologies of the 1990s and allowed web users to interact with websites and contribute their own content to the websites. Social media embodied the principles of Web 2.0 (O'Reilly, 2007) and created a paradigm shift in the way organizations operate and collaborate. As social media is tremendously popular among consumers, firms can leverage it to engage in frequent and direct consumer contact with a broad reach at a relatively low cost (Kaplan & Haenlein, 2010).

Social media analytics support social media content mining, usage mining, and structure mining activities. Social media analytics analyze and interpret human behaviors at social media sites, providing insights and drawing conclusions from a consumer's interests, web browsing patterns, friend lists, sentiments, profession, and opinions. By understanding customers better using social media analytics, firms develop effective relationship marketing campaigns for targeted customer segments and tailor products and services to customers' needs and interests. For example, major U.S. banks analyze clients' comments on social media sites about their service experiences and satisfaction levels. Unlike web analytics used mainly for structured data, social media analytics are used for the analysis of data likely to be natural language, unstructured, and context-dependent.

The worldwide social media analytics market is growing rapidly from $1.6 billion in 2015 to an estimated $5.4 billion by 2020 at a compound annual growth rate of 27.6%. This growth is attributable to advanced analytics and the increase in the number of social media users (ReportsnReports, 2016). Some social media analytics software programs are provided as cloud-based services with flexible fee options, such as monthly subscription or pay-as-you-go pricing. Social media analytics focus on two types of analysis: sentiment analysis and social network analysis.

Sentiment analysis uses text analysis, natural language processing, and computational linguistics to identify and extract user sentiments or opinions from text materials. Sentiment analysis can be performed at multiple levels, such as entity level, sentence level, and document level. An entity-level analysis identifies and analyzes individual entity's opinions contained in a document. A sentence-level analysis identifies and analyzes sentiments expressed in sentences. A document-level analysis identifies and analyzes an overarching sentiment expressed in the entire document. While a deeper understanding of sentiment and greater accuracy are still to be desired, sentiments extracted from documents have been used successfully by businesses in various ways, including predicting stock market movements, determining market trends, analyzing product defects, and managing crises (Fan & Gordon, 2014). However, sentiment analysis can be flawed. Sampling biases in the data can skew results as in situations where satisfied customers remain silent while those with more extreme positions express their opinions (Fan & Gordon, 2014).

Lexical-based methods and machine-learning methods are two widely used methods for sentiment analysis. Lexical-based methods use a predefined set of words in which each word carries a specific sentiment. These methods include:

- simple word or phrase counts;
- the use of emoticons to detect polarity; that is, positive and negative emoticons used in a message (Park, Barash, Fink, & Cha, 2013);
- sentiment lexicons, based on the words in the lexicon that have received specific features marking the positive or negative terms in a message (Gayo-Avello, 2011); and
- the use of psychometric scales to identify mood-based sentiments.

One of the challenges of the lexical-based methods is to create a lexical-based dictionary to be used for different contexts.

Machine-learning methods often rely on the use of supervised and unsupervised machine-centered techniques. While one advantage of machine-learning methods is the ability to adapt and generate trained models for specific purposes and contexts, the drawback to these methods is the lacking availability of labeled data and hence the low applicability of the methods to new situations (Gonçalves, Araújo, Benevenuto, & Cha, 2013). In addition, labeling data might be costly or even prohibitive for some tasks. While machine-learning methods are reported to perform better than lexical-based methods, it is hard to conclude whether a single machine-learning method is better than all lexical-based methods across different tasks.

Social network analysis is the process of measuring the social network structure, connections, nodes, and other properties by modeling social network dynamics and growth (e.g., network density, network centrality, network flows). Social network analysis originally was developed before the advent of social media to study relationships among actors in modern sociology. The relationships typically are identified from links directly connecting two actors or inferred indirectly from tagging, social-oriented interactions, content sharing, and voting. Social networking sites such as Facebook, Twitter, and LinkedIn provide a central point of access and bring structure in the process of personal information sharing and online socialization (Jamali & Abolhassani, 2006).

Social network analysis uses a variety of techniques pertinent to understanding the structure of the network (Scott, 2012). These techniques range from simpler methods, such as counting the number of edges a node has or computing path lengths, to more sophisticated methods that compute eigenvectors to determine key nodes in a network (Fan & Gordon, 2014). Social networking sites have been a popular subject of social network analysis. Marlow (2004) employs social network analysis to describe the social structure of blogs. He explores two metrics of authority: popularity measured by bloggers' public affiliations and influence measured by citation of the writing.

### 4.3. Big Data 3.0 (2015—)

Big Data 3.0 encompasses data from Big Data 1.0 and Big Data 2.0. The main contributors of Big Data 3.0 are the IoT applications that generate data in the form of images, audio, and video. The IoT refers to a technology environment in which devices and sensors have unique identifiers with the ability to share data and collaborate over the internet even without any human intervention. With the rapid growth of the IoT, connected devices and sensors will surpass social media and e-commerce websites as the primary sources of big data. GE is developing IoT-based sensors that read data from equipment deployed for aviation and healthcare operations. Agribusinesses also use IoT-based sensors to manage resources like water, grain storage, and heavy equipment in an effort to drive down agricultural costs and increase yields.

For many IoT applications, the analysis increasingly is performed by sensors at the source of data gathering. This trend is leading to a new field known as streaming analytics. Streaming analytics continuously extract information from the streaming data. Unlike social media analytics used in a batch mode for the analysis of stored data, streaming analytics involve real-time event analysis to discover patterns of interest as data is being collected or generated. Streaming analytics are used not just for monitoring existing conditions but also for predicting future events.

Streaming analytics have great potential in a number of industries where streaming data are generated through human activities, machine data, or sensor data. For example, streaming analytics embedded in sensors can monitor and interpret patients' physiological and behavioral changes and alert caregivers to urgent medical needs. Streaming analytics also can be useful in the financial industry where electronic transactions need to be monitored under financial regulations and immediate actions are required in the event of suspicious and fraudulent financial activities.

## 5. An illustrative example: Big data analysis of merchant reviews

With the explosive growth of merchant reviews at various vendor/product review sites and social media sites, merchant review big data has drawn the attention of researchers and practitioners. Reviews written by consumers are perceived to be less biased than those provided by advertisers or product experts. The review credibility can be further enhanced by providing a feedback function for viewers to rate the usefulness of the particular reviews. Yelp, TripAdvisor, and Angie's List are popular merchant review sites. These sites enable consumers to rate a particular merchant based on a numerical scale (e.g., 1 to 5). They provide viewers with the entirety of consumers' reviews along with the ability to vote on the helpfulness of those reviews. It would be challenging for small business merchants to analyze their review data, since the data are large and unstructured. Despite the popularity of the merchant review, it is still the case that merchants fail to fully exploit and translate consumer reviews into business value.

This section illustrates a simple but powerful application of social media analytics to merchant review data. A social media analytics model was developed to discover relationships between consumers' review activities and the viewers' usefulness votes in the context of Groupon users' merchant reviews. Five factors related to consumers' review activities were identified that may influence the usefulness of the review. The five factors include (1) the review score of the reviewer, (2) the number of social network friends of the reviewer, (3) the cumulative number of reviews made by the reviewer, (4) the number of words in the reviewer's comment, and (5) the existence of images or photos in the reviewer's comment. Note that the number of words is derived from the reviewer's comment, which is originally in an unstructured form. The existence of images or photos is a dummy variable in this model. The dependent variable is the number of usefulness votes by viewers.

A multiple regression model is used to identify the factors that are strongly associated with the number of usefulness votes. Merchant reviews were collected in July 2015 from 108 healthcare merchants that launched Groupon promotions between June and July 2011. The term healthcare, as used by Groupon, refers to a variety of businesses and services, from a haircut salon and spa to fitness or yoga training. Groupon user reviews were identified in July 2015 on the 108 healthcare merchant listings on Yelp and these reviews were transcribed for analysis. Out of 589 reviews, 189 reviews were removed that did not have any response from viewers, leaving 400 reviews that were analyzed using a multiple regression model. The descriptive statistics are shown in Table 1. Table 2 shows the beta coefficients of these variables as well as their p-values. The regression was run with a 5% level of significance.

The results show that the review score, the number of social network friends of a reviewer, and the number of words in a review are significant predictors of the number of usefulness votes. The cumulative number of reviews made by a reviewer and the existence of images/photos do not have an impact on the number of usefulness votes. It is interesting to note that as the review score increases, the number of usefulness votes decreases. The review score's negative effect on the number of

**Table 1. Descriptive statistics of merchant review**

| | *Mean* | *Standard Deviation* | *n* |
|---|---|---|---|
| *Review Score* | 3.365 | 1.5705 | 400 |
| *Number of Friends* | 72.0675 | 165.3403 | 400 |
| *Number of Reviews* | 96.37 | 168.382 | 400 |
| *Number of Words* | 221.96250 | 163.2673 | 400 |
| *Image/Photo* | 0.0375 | 0.1899 | 400 |

usefulness votes indicates that when review scores are low, viewers feel the review is more useful for their purchase decision making. The reviewers who have more friends in the social network influence the viewers' votes. Therefore, merchants need to pay more attention to network leaders with higher numbers of social network friends. This result is consistent with a finding that consumers are more likely to trust a reviewer who has a higher number of followers (Cheung & Ho, 2015). Finally, the number of words in the reviewer's comment has a significant positive effect on the number of usefulness votes by viewers. This result may be explained by the fact that the higher number of words brings more helpful information to viewers and leads to reduced information asymmetry between the merchant and the viewers. While more comprehensive social media analytics might add more value to merchants, this illustration shows that even simple data analytics can deliver highly valuable marketing ideas.

## 6. Impacts of big data

Big data provides great potential for firms in creating new businesses, developing new products and services, and improving business operations. The

**Table 2. Results of the regression**

| *Multiple Linear Regression - Estimated Regression Equation* | | | | | |
|---|---|---|---|---|---|
| useful[t] = +3.18935 − 0.632662score[t] + 0.00557255friend[t] + 0.00135261review[t] + 0.00921948word[t] + 0.717262image[t] + 0.00192933t + e[t] | | | | | |
| *Multiple Linear Regression - Ordinary Least Squares* | | | | | |
| Variable | Parameter | S.D. | T-STAT H0: parameter = 0 | 2-tail p-value | 1-tail p-value |
| (Intercept) | +3.189 | 0.8144 | +3.9160e+00 | 0.0001061 | 5.306e-05 |
| score | −0.6327 | 0.1469 | −4.3080e+00 | 2.083e-05 | 1.041e-05 |
| friend | +0.005573 | 0.001739 | +3.2050e+00 | 0.001461 | 0.0007305 |
| review | +0.001353 | 0.001692 | +7.9920e-01 | 0.4247 | 0.2123 |
| word | +0.00922 | 0.001384 | +6.6610e+00 | 9.169e-11 | 4.585e-11 |
| image | +0.7173 | 1.203 | +5.9620e-01 | 0.5514 | 0.2757 |
| t | +0.001929 | 0.001976 | +9.7620e-01 | 0.3296 | 0.1648 |
| *Multiple Linear Regression - Regression Statistics* | | | | | |
| Multiple R | | | | | 0.4443 |
| R-squared | | | | | 0.1974 |
| Adjusted R-squared | | | | | 0.1852 |
| F-TEST (value) | | | | | 16.11 |
| F-TEST (DF numerator) | | | | | 6 |
| F-TEST (DF denominator) | | | | | 393 |
| p-value | | | | | 1.11e-16 |
| *Multiple Linear Regression - Residual Statistics* | | | | | |
| Residual Standard Deviation | | | | | 4.437 |
| Sum Squared Residuals | | | | | 7737 |

use of big data analytics can create benefits, such as cost savings, better decision making, and higher product and service quality (Davenport, 2014). Personalized advertising that is finely tuned to what consumers are looking for and news articles related to their interests are some of the impacts of big data (Goetz, 2014). It is also noted that while managers realize that big data has potential impacts on firms, they still face difficulty in exploiting the data. The following discusses the impacts of big data on large firms.

### 6.1. Personalization marketing

By exploiting big data from multiple sources, firms can deliver personalized product/service recommendations, coupons, and other promotional offers. Major retailers such as Macy's and Target use big data to analyze shoppers' preferences and sentiments and improve their shopping experience. Innovative fintech firms have already started using social media data to assess the credit risk and financing needs of potential clients and provide new types of financial products for them. Banks are analyzing big data to increase revenue, boost retention of clients, and serve clients better. U.S. Bank, a major commercial bank in the U.S., deployed data analytics that integrate data from online and offline channels and provide a unified view of clients to enhance customer relation management. As a result, the bank's lead conversion rate has improved by over 100% and clients have been able to receive more personalized experiences (The Financial Brand, 2014).

### 6.2. Better pricing

Harnessing big data collected from customer interactions allows firms to price appropriately and reap the rewards (Baker, Kiewell, & Winkler, 2014). Sears uses big data to help set prices and give loyalty shoppers customized coupons. Sears deployed one of the largest Hadoop clusters in the retail industry and now utilizes open source technologies to keep the cost of big data low. Sears analyzes massive amounts of data about product availability in its stores to prices at other retailers to local weather conditions in order to set prices dynamically. eBay also uses open source Hadoop technology and data analytics to optimize prices and customer satisfaction. To achieve the highest price possible for items sellers place for auction, eBay examines all data related to items sold before (e.g., a relationship between video quality of auction items and bidding prices) and suggests ways to maximize results to sellers.

### 6.3. Cost reduction

Big data reduces operational costs for many firms. According to Accenture (2016), firms that use data analytics in their operations have faster and more effective reaction time to supply chain issues than those that use data analytics on an ad-hoc basis (47% vs. 18%). Big data analytics leads to better demand forecasts, more efficient routing with visualization and real-time tracking during shipments, and highly optimized distribution network management (House, 2014).

GE helps the oil and gas industry improve equipment reliability and availability, resulting in better operations efficiency and higher oil and gas productivity. Real-time monitoring systems transmit massive amounts of data to central facilities where they are processed with data analytics to assess equipment conditions. GE provides Southwest Airlines with proprietary flight efficiency analytics to analyze flight and operational data in order to identify and prioritize fuel savings opportunities (Business Wire, 2015).

Big data has also led to enormous cost reduction in the retail industry. Tesco, a European supermarket store, analyzes refrigerator data to reduce energy cost by about $25 million annually. Analyses of refrigerator data showed that the temperature of refrigerators were set colder than necessary, wasting electricity. To optimize the temperature of the refrigerators, all Tesco refrigerators in Ireland were equipped with sensors that monitored the temperature every 3 seconds (van Rijmenam, 2016).

### 6.4. Improved customer service

Big data analytics can integrate data from multiple communication channels (e.g., phone, email, instant message) and assist customer service personnel in understanding the context of customer problems holistically and addressing problems quickly. Big data analytics can also be used to analyze transaction activities in real time, detect fraudulent activities, and notify clients of potential issues promptly. Insurance claim representatives can serve clients proactively based on the correlation analysis of weather data and certain types of claims submitted on stormy or snowy days.

Hertz, a car rental company in the U.S., uses big data to improve customer satisfaction. Hertz gathers data on its customers from emails, text messages, and online surveys to drive operational improvement. For example, Hertz discovered that return delays were occurring during specific hours of a day at an office in Philadelphia and was able to add staff during the peak activity hours to make sure that any issues were resolved promptly (IBM, 2010). Southwest Airlines uses speech analytics to extract business intelligence from conversations between customers and the company's service personnel. The airline also uses social media analytics to delve into customers' social media data for a better understanding of customer intent and better service offerings (Aspect, 2013).

## 7. Challenges in big data

Based on the survey of big data practices, in this section I discuss challenges in big data development and management. As with any disruptive innovation, big data presents multiple challenges to adopting firms. For example, SAS (2013) notes that enterprises will face challenges in processing speed, data interpretation, data quality, visualization, and exception handling of big data. I highlight six technical and managerial challenges: data quality, data security, privacy, investment justification, data management, and shortage of qualified data scientists.

### 7.1. Data quality

Data quality refers to the fitness of data with respect to a specific purpose of usage. Data quality is critical to confidence in decision making. As data are more unstructured and collected from a wider array of sources, the quality of data tends to decline. For firms adopting data analytics for their supply chain, data quality is paramount. If the data are not of high quality, managers will not use the data, let alone want to share the data with their partners. Streaming analytics use data generated by interconnected sensors and communication devices. If a medical monitoring system's sensor generates erroneous data, the streaming analytics may send a wrong signal to the controlling devices that may be fatal to patients. A data quality control process needs to be established to develop quality metrics, evaluate data quality, repair erroneous data, and assess a trade-off between quality assurance costs and gains.

### 7.2. Data security

Weak security creates user resistance to the adoption of big data. It also leads to financial loss and damage to a firm's reputation. Without installing proper security mechanisms, confidential information could be transmitted inadvertently to unintended parties. This security challenge may be alleviated by establishing strong security management protocol, along with security solutions such as intrusion prevention and detection systems, encryptions, and firewalls built into big data systems. Blockchain, an underlying technology behind the Bitcoin cryptocurrency, is a promising future technology for big data security management. Saving data in an encrypted form rather than in its original format, blockchain ensures that each data element is unique, time-stamped, and tamper-resistant, and its applications extend beyond financial industries due to the enhanced level of data security.

### 7.3. Privacy

As big data technologies mature, the extensive collection of personal data raises serious concerns for individuals, firms, and governments. Without addressing these concerns, individuals may find data analytics worrisome and decide not to contribute personal data that can be analyzed later. According to the 2015 TRUSTe Internet of Things Privacy Index, only 20% of online users believe that the benefits of smart devices outweighed any privacy concerns (TRUSTe, 2015). As is the case with smart health equipment and smart car emergency services, sensors can provide a vast amount of data on users' location and movements, health conditions, and purchasing preferences, all of which raise significant privacy concerns. However, protecting privacy is often counterproductive to both firms and customers, as big data is a key to enhanced service quality and cost reduction. Therefore, firms and customers need to strike a balance between the use of personal data for services and privacy concerns. It is noted that there is no one-size-fits-all measure for privacy, but the balance depends on service type, customers served, data type, and regulatory environments.

### 7.4. Investment justification

According to Accenture (2016), the actual use of big data analytics is limited. Despite the touted benefits of big data, firms face difficulties in proving the value of big data investments. A majority of surveyed executives (67%) expressed concerns about the large investment required to implement and use analytics. Many big data projects have unclear problem definitions and use emerging technologies, thus causing a higher risk of project failure and higher irreversibility of investments than traditional technology projects. In addition, if tangible costs significantly outweigh tangible benefits—despite potentially large intangible benefits—it will be hard to justify investment to senior management due to the calculated negative financial returns. When a project is highly risky and irreversible, a real option approach may be appropriate (Lee & Lee, 2015). In a real option approach, options such as postponement, expansion, shrinkage, and scrapping of a project are viable, and there is not an obligation to move forward with a big data plan as-is.

### 7.5. Data management

Social media and streaming sensors generate massive amounts of data that need to be processed. Few firms would be able to invest in data storage for all big data collected from their sources. The current architecture of the data center is not prepared to deal with the heterogeneous nature of personal and enterprise data (Gartner, 2015). Deutsche Bank has been working on big data implementation since the beginning of 2012 in an attempt to analyze all of its unstructured data. However, problems have occurred when trying to make big data applications work with its traditional mainframes and databases. Petabytes of data had been stored across dozens of data warehouses, but extracting these for analysis became an expensive proposition (The Financial Brand, 2014).

A combination of edge computing and Hadoop has the potential to help reduce data management issues for firms. Hadoop is useful for complex transformations and computations of big data in a distributed computing environment. However, Hadoop is not suitable for ad hoc data exploration and streaming analytics. The need for streaming analytics and real time responses is driving the development of edge computing, also known as fog computing. However, edge computing is costlier to develop and maintain than data centers.

### 7.6. Shortage of qualified data scientists

As the need to manipulate unstructured data such as text, video, and images increases rapidly, the need for more competent data scientists grows. According to an A.T. Kearney survey of 430 senior executives, despite the prediction that firms will need 33% more big data specialists over the next 5 years, roughly 66% of firms with advanced analytics capabilities were not able obtain enough employees to deliver insights into their big data (Boulton, 2015). The McKinsey Global Institute estimated that the U.S. needs 140,000—190,000 more workers with analytical skills and 1.5 million managers and analysts with analytical skills to make business decisions based on the analysis of big data (Manyika et al., 2011). IDC (2015) reported that the staff shortage will extend from data scientists to data architects and experts in data management, and big data-related professional services will grow at a compound annual rate of 23% through 2020 (Vesset et al., 2015). If this shortage continues, firms will have to offer highly competitive salaries to qualified data scientists and may need to develop data analytics training programs in order to groom internal employees to meet the demand.

## 8. The future of big data

Big data's emergence has not remained isolated to a few sectors or spheres of technology, instead demonstrating broad applications across industries. In light of this reality, companies must first pursue big data capabilities as necessary ground-level developments, which in turn may facilitate competitive advantages. Formidable challenges face firms in pursuit of big data integration, but the potential benefits of big data promise to positively impact company operations, marketing, customer experience, and more. Using tools to assess and understand big data, such as the integrated view of big data dimensions presented here (see Figure 1), will help companies both to realize the individual benefits of big data integration and to position themselves within the wider technological shift as big data becomes a part of mainstream business practices. There is a need for more practical research examining and tackling these challenges to big data within business, as well as a need for industry changes to encourage talent and infrastructure development. This big data reality in which companies find themselves is vast, complex, comprehensive—and here to stay.

## References


Accenture. (2016). *Big data analytics in supply chain: Hype or here to stay?* Available at https://www.accenture.com

Aspect. (2013, August 27). *Southwest Airlines heads to the cloud with Aspect software to provide best-in-class customer experience* [Press Release]. Retrieved from http://www.aspect.com/company/news/press-releases/southwest-airlines-heads-to-the-cloud-with-aspect-software-to-provide-best-in-class-customer-experience

Baker, W., Kiewell, D., & Winkler, G. (2014). Using big data to make better pricing decisions. *McKinsey*. Retrieved from http://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/using-big-data-to-make-better-pricing-decisions

Boulton, C. (2015). Lack of big data talent hampers corporate analytics. *CIO*. Retrieved from http://www.cio.com/article/3013566/analytics/lack-of-big-data-talent-hampers-corporate-analytics.html

Business Wire. (2015, September 30). *Southwest Airlines selects GE's Flight Efficiency Analytics* [Press Release].

Retrieved from http://www.businesswire.com/news/home/20150930006279/en/

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly, 36*(4), 1165–1188.

Cheung, Y.-H., & Ho, H.-Y. (2015). Social influence's impact on reader perceptions of online reviews. *Journal of Business Research, 68*(4), 883–887.

Davenport, T. H. (2014). *Big data at work: Dispelling the myths, uncovering the opportunities*. Boston: Harvard Business Review Press.

Fan, W., & Gordon, M. D. (2014). The power of social media analytics. *Communications of the ACM, 57*(6), 74–81.

The Financial Brand. (2014). *Big data: Profitability, potential, and problems in banking*. Retrieved from http://thefinancialbrand.com/38801/big-data-profitability-strategy-analytics-banking/

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*(2), 137–144.

Gartner. (2015, November 10). *Gartner says 6.4 billion connected things will be in use in 2016, up 30 percent from 2015*. Retrieved from http://www.gartner.com/newsroom/id/3165317

Gayo-Avello, D. (2011). Don't turn social media into another 'Literary Digest' poll. *Communications of the ACM, 54*(10), 121–128.

Goetz, H. (2014, March 26). What Google now can teach enterprises about big data. *Prolifiq*. Retrieved from http://prolifiq.com/google-now-big-data/

Gonçalves, P., Araújo, M., Benevenuto, F., & Cha, M. (2013). Comparing and combining sentiment analysis methods. In *Proceedings of the 2013 Conference on Online Social Networks* (pp. 27–37). Boston: ACM.

House, J. (2014, November 20). Big data analytics = Key to successful 2015 supply chain strategy. *ModusLink*. Retrieved from https://www.moduslink.com/big-data-analytics-key-successful-2015-supply-chain-strategy/

IBM. (2010). *How big data is giving Hertz a big advantage*. Retrieved from https://www-01.ibm.com/software/ebusiness/jstart/portfolio/hertzCaseStudy.pdf

IDC. (2015). *New IDC forecast sees worldwide big data technology and services market growing to $48.6 billion in 2019, driven by wide adoption across industries* [Press Release]. Retrieved from http://www.idc.com/getdoc.jsp?containerId=prUS40560115

Jamali, M., & Abolhassani, H. (2006). Different aspects of social network analysis. In *Proceedings — 2006 IEEE/WIC/ACM International Conference on Web Intelligence* (pp. 66–72). Hong Kong: IEEE.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media *Business Horizons, 53*(1), 59–68.

Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience, and acquisition intention of big data analytics. *International Journal of Information Management, 34*(3), 387–394.

Laney, D. (2001, February 6). 3D data management: Controlling data volume, velocity, and variety. *META Group*. Retrieved from http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Lee, I., & Lee, K. (2015). The Internet of Things (IoT): Applications, investments, and challenges for enterprises. *Business Horizons, 58*(4), 431–440.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., et al. (2011, May). Big data: The next frontier for innovation, competition, and productivity. *McKinsey*. Available at http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation

Marlow, C. (2004). Audience, structure, and authority in the weblog community. In *Proceedings of the 54th Annual Conference of the International Communication Association*. Available at http://alumni.media.mit.edu/~cameron/cv/pubs/04-01.pdf

O'Reilly, T. (2007). What is Web 2.0? Design patterns and business models for the next generation of software. *Communications and Strategies, 65*, 17–37.

Park, J., Barash, V., Fink, C., & Cha, M. (2013). Emoticon style: Interpreting differences in emoticons across cultures. In *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media* (pp. 466–475). Boston: AAAI.

ReportsnReports. (2016, February 12). Social media analytics market to rise at 27.6% CAGR to 2020. *PRNewswire*. Retrieved from http://www.prnewswire.com/news-releases/social-media-analytics-market-to-rise-at-276-cagr-to-2020-568584751.html

SAS. (2013). *Five big data challenges*. Retrieved from https://www.sas.com/resources/asset/five-big-data-challenges-article.pdf

Scott, J. (2012). *Social network analysis* (3rd ed.). Thousand Oaks, CA: Sage.

TRUSTe. (2015). *2015 US IoT Privacy Index [Infographic]*. Retrieved from https://www.truste.com/resources/privacy-research/us-internet-of-things-index-2015/

van Rijmenam. (2016). Tesco and big data analytics, a recipe for success? *Datafloq*. Retrieved from https://datafloq.com/read/tesco-big-data-analytics-recipe-success/665

Vesset, D., Olofson, C. W., Nadkorni, A., Zaidi, A., Mcdonough, B., Schubmehl, D., et al. (2015). Futurescape: Worldwide big data and analytics 2016 predictions. *IDC*. Retrieved from https://www.idc.com/research/viewtoc.jsp?containerId=259835

*Source: Sample Article. Sourced from "Big data : Dimensions , evolution , impacts , and challenges," by I. Lee, 2017, Business Horizons, 60(3), 293–303. https://doi.org/10.1016/j.bushor.2017.01.004*

## 6. Appendix C

Introduction

**As part of my MBA journey at the Gordon Institute of Business Science (GIBS), I am conducting research into the enabling factors for formulating an effective data led strategy (EDLS) in organisations with a big data and analytics capability or intention thereof. As such, I would highly appreciate your assistance in completing my questionnaire to fulfil my research requirements. This questionnaire will take between 10 to 12 minutes to complete. All responses will remain anonymous. All data that is collected will only be reported at an aggregated level. By completing this questionnaire, you are indicating that your participation is voluntary and you are able to exit the questionnaire at any point if required.**

**For any further questions or details, please feel free to contact myself or my supervisor on the details below.**

**Researcher: Gavin Reuben**

**Contact Details: 18361562@mygibs.co.za or +2771 672 0351**

**Supervisor: Dr Jeff Chen**

**Contact Details: chenj@gibs.co.za**

Part A: Descriptive. Please choose the appropriate answer.

1. I am aware of the integration of big data and relevant data insights to generate key strategies?

- ☐ Yes
- ☐ No

2. I am...

- ☐ in a big data and analytics team
- ☐ a consumer of data driven insights
- ☐ Not applicable

3. Has there been a positive impact in your organisations from your data capabilities and strategy?

- ☐ Yes
- ☐ No

4. My organisation size can be best described as ...

- ☐ Small (< 51 employees)
- ☐ Medium to Large (> 50 employees)

5. Please select an industry best suited to your organisation

6. Please select the best description of your current job level

Part B: Indicate the extent of the following where 1 equals strongly disagree and 5 strongly agree.

**In the context of a big data and strategy, please provide the most accurate response based on your organisation and experience.**

7. Compared to rivals within my industry, my organisation has the foremost available analytics systems to deliver strategic insights by leveraging big data.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

8. The legacy systems in my organisation restricts the development of new big data applications for strategy formulation.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

9. The big data infrastructure in my organisation can enable effective strategic insights through analytics.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

10. Our team has explored or adopted recent big data technologies and approaches to data processing.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

11. Our big data analytics staff has the required technical skills to accomplish their jobs successfully.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

12. Our big data analytics staff has suitable formal qualification to fulfil their jobs.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

13. Our big data analytics staff holds suitable work experience to accomplish their jobs successfully.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

14. We provide big data analytics training to our own employees.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

15. My organisation has access to very large, unstructured, or fast-moving data for analysis to derive insights.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

16. My organisation integrates data from multiple internal sources into a data warehouse, enabling the ability for strategic decision making.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

17. My organisation integrates external data sets with internally generated data to facilitate high-value analysis of our strategic direction.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

18. Ensuring the validity and veracity (accuracy) of our data is a key concern for my organisation.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

19. Decisions in my organisation are largely based on data rather than and individuals perception.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

20. My organisation continuously improves its business strategy as a result of insights extracted from big data.

- ◯ Strongly agree
- ◯ Agree
- ◯ Neither agree nor disagree
- ◯ Disagree
- ◯ Strongly disagree

21. I am willing to override my own intuition when data contradicts my viewpoints.

- ◯ Strongly agree
- ◯ Agree
- ◯ Neither agree nor disagree
- ◯ Disagree
- ◯ Strongly disagree

22. All employees in my organisation are encouraged to make strategic decisions based on big data.

- ◯ Strongly agree
- ◯ Agree
- ◯ Neither agree nor disagree
- ◯ Disagree
- ◯ Strongly disagree

We almost at the halfway mark...

**In the context of a big data and strategy, please provide the most accurate response based on your organisation and experience.**

23. My organisation highly values the attitudes that promote internal changes.

- ◯ Strongly agree
- ◯ Agree
- ◯ Neither agree nor disagree
- ◯ Disagree
- ◯ Strongly disagree

24. In my organisation we collaborate with different institutions in innovation projects.

- ◯ Strongly agree
- ◯ Agree
- ◯ Neither agree nor disagree
- ◯ Disagree
- ◯ Strongly disagree

25. My organisation constantly seeks relevant external information for its business.

- ◯ Strongly agree
- ◯ Agree
- ◯ Neither agree nor disagree
- ◯ Disagree
- ◯ Strongly disagree

26. My organisation heavily invests in research and development activities.

- ◯ Strongly agree
- ◯ Agree
- ◯ Neither agree nor disagree
- ◯ Disagree
- ◯ Strongly disagree

27. My organisation often uses knowledge or technology developed by other companies.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

28. Managers have appropriate knowledge for the development of their functions.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

29. Employees are often encouraged to participate in scientific events (conferences, seminars, courses).

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

30. My organisation promotes integration and sharing of knowledge between different sectors.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

31. My organisation promotes exchange of experience and knowledge between its divisions.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

32. Information moves with ease and agility among the hierarchical levels of the organisation.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

33. My organisation usually practices the turnover of roles and tasks among employees.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

34. There is spontaneous cooperation among the employees at all levels.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

35. My organisation applies its accumulated knowledge to develop technology strategy.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

36. My organisation is capable of incorporating technological knowledge in patents.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

37. My organisation responds nimbly to business changes using new knowledge.

- ◯ Strongly agree
- ◯ Agree
- ◯ Neither agree nor disagree
- ◯ Disagree
- ◯ Strongly disagree

38. My organisation seeks to innovate ahead of its competitors.

- ◯ Strongly agree
- ◯ Agree
- ◯ Neither agree nor disagree
- ◯ Disagree
- ◯ Strongly disagree

## Almost done...

**In the context of a big data and strategy, please provide the most accurate response based on your organisation and experience.**

39. For decisions with a relatively high level of perceived impact, it is more important to reduce the danger to overlook potentially relevant issues than to enhance the speed and accuracy of decision-makers´ action.

- ◯ Strongly agree
- ◯ Agree
- ◯ Neither agree nor disagree
- ◯ Disagree
- ◯ Strongly disagree

40. In my organisation when ambiguity surrounds issues and action alternatives is high, it leads to higher innovation performance.

- ◯ Strongly agree
- ◯ Agree
- ◯ Neither agree nor disagree
- ◯ Disagree
- ◯ Strongly disagree

41. There is disbelief in my organisation on the potential impact of data enabling strategy, to create value.

- ◯ Strongly agree
- ◯ Agree
- ◯ Neither agree nor disagree
- ◯ Disagree
- ◯ Strongly disagree

42. There is an over emphasis in my organisation on the obvious and tangible features within the environment of data led strategy.

- ◯ Strongly agree
- ◯ Agree
- ◯ Neither agree nor disagree
- ◯ Disagree
- ◯ Strongly disagree

43. In my organisation, in situations of relatively high uncertainty, there is a possibility that we overlook potentially relevant issues and answers.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

44. In my organisation, in situations of relatively low uncertainty, regulating the organisations decision-makers' attention with inflexible structures improves the speed and accuracy of their actions.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

45. In my organisation, in situations of relatively low uncertainty, enhancing the speed and accuracy of decision-makers´ action is favourable.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

46. In my organisation, in situations of relatively high uncertainty, reducing the possibility to overlook potentially relevant issues and answers is preferable.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

47. In my organisation, there is an alignment between corporate and divisional attention to threats and opportunities.

- ◯ Strongly agree
- ◯ Agree
- ◯ Neither agree nor disagree
- ◯ Disagree
- ◯ Strongly disagree

48. In my organisation there is a focus of corporate and business unit attention on strategic (competitive and long-term) issues, which facilitate business unit adaptation (response to threats and opportunities).

- ◯ Strongly agree
- ◯ Agree
- ◯ Neither agree nor disagree
- ◯ Disagree
- ◯ Strongly disagree

49. In my organisation there is a coordination of business unit adaptation by linking attention to ongoing planning, financial, operational and human resource issues and initiatives.

- ◯ Strongly agree
- ◯ Agree
- ◯ Neither agree nor disagree
- ◯ Disagree
- ◯ Strongly disagree

50. In my organisation there is a tight coupling of governance channels that ensure adaptive moves will be coordinated vertically across levels of the organisation and horizontally across its various functions and channels.

- ◯ Strongly agree
- ◯ Agree
- ◯ Neither agree nor disagree
- ◯ Disagree
- ◯ Strongly disagree