Is Big Data too Big for SMEs?

MS&E-238: Leading Trends in Information Technology
Stanford University
Summer 2014

Carl Johan Rising - SUID: 006000731
Michael Kristensen - SUID: 005999170
Steffen Tjerrild-Hansen - SUID: 005997833

Table of content

1. Introduction

The increasing focus on big data and how it has potential to influence almost every industry, gives it a often deterministic presence that presents it as a implement-and-reap beneficial solution for enterprises. However, many organizations and often SMEs fail to successfully implement technological and organizational frameworks in order to build the capabilities to harness some of the potential that data, both small and big, can hold. The early uses of gathering a large variety and volume of data has mostly been within large corporations and used for fraud detection and coupon systems for retailers. The project is inspired by the limited focus on the potential for SMEs to harness big data instead of multinational enterprises. As noted, mostly larger enterprises has launched initiatives to compliment their analytical proficiencies, but as technologies mature, and more companies adopt frameworks for handling data, and learn how to organize within this new framework, SMEs might find an easier time reaping some of the benefits. Also led by cheaper and more easily accessible servers and data centers, delivered through cloud vendors, SMEs now face less of a constraint on upfront investment, rather the challenges presents themselves as organizational and strategic of nature. The right technologies still needs to be chosen, but with well supported and documented open source data systems being available, it has increasingly become a question of choosing right, and choosing a scalable option that fits the specific need of an SME.

2. Research question

The main objective of the project is to investigate the following research question:

How might SMEs build their technological infrastructure and set up their organization in order to derive value from big data to strengthen their strategy formation?

In order to answer and provide valid recommendations for SMEs are present academic literature and other relevant sources examined.

2.1 Scope

First of all is it important to state that building the right infrastructure and organization in order to harness big data strategically is very different from industry to industry and from company to company. This project will mainly focus on big data within SMEs on a general level.

In order to provide a holistic and tangible action plan for SMEs to harness big data is the project focusing on three distinct areas; technology, organization and strategy. We look into what infrastructure SMEs should put in place to take advantage of big data capabilities in strategy formation. This is done while consistently drawing comparison to current publications and organizations.

The organizational section of the project will mainly focus on how the path to big data might differ for SMEs compared to the ones taken by big enterprises. More specifically, it will deal with how necessary skills are attained, how SMEs organize profiles with these skills and what should accompany in the organization to secure value, i.e. Information Governance. The section will not comment directly on what impact all these actions may have on other departments.

The strategical part will mainly focus on the high level potential and how data can influence performance of SMEs. It's hard to generalize concrete strategic recommendation for SMEs because its so specific to the individual company. Only special strategical use cases will be touched upon, such as how personalized service is being reinvented because of big data.

2.2 Structure

In order to answer this question, we first need to give a definition on big data in this setting and through our own understanding. This goes in accordance to the three first concerns addressed by top IT and business leaders in Gartner's Research Circle (Kart, Heudecker & Buytendijk, 2013). Next, will the technological capabilities and the optimal infrastructure of enterprises be elaborated. The next section will focus on the optimal organisational setup with in SMEs in order to utilize the potential of big data. This leads to

the strategical section where the potential of big data within SMEs will be investigated. Finally will we take a step back and briefly touch why many big data projects often fail within SMEs and instead how to successfully implement big data.

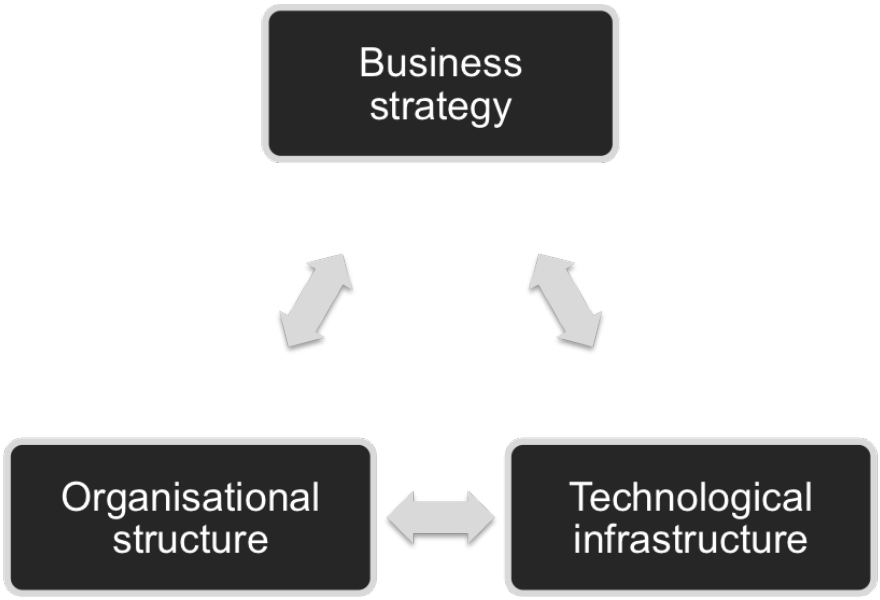


Figure 1: (Pearlson & Saunders, 2006)

2.3 Limitations

It is important to distinct between what kind of workload requirements there might be for a system, we are in this project mostly concerned with systems supporting decision making processeses. This is not however the full scope of the capabilities that might arise from these systems, as we are aware, but may also include workload and document handling intensive systems.

Further, this paper can be considered to draw tentative conclusions, as Big Data within SMEs is a limited researched area in which use cases in academic literature are scarce - although on their way ("EIT ICT Labs," n.d.). Thus, most of the conclusions draw in this paper are based on use cases described in business analyst publications, news articles and appropriated conclusions from related research areas.

3. Defining big data and conceptual clarification

Big data is undoubtedly one of the hottest topics in computing and yet it seems like nobody has the same definition of the term. Different variations are used in academia, industry and the media, and numerous stakeholders provide diverse and often contradictory

definitions. Therefore, it is important to provide clear definition of big data in order to align the reader's individual perception of the term.

After reviewing major influential organizations different definitions (MIT Tech Review, 2013) of the term was Gartner's version found the most appropriate for this project.

"Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision-making." (Gartner, 2014)

Gartner's definition is broader and encompasses many different aspects of big data, which is deviating from more literal versions as NIST's (The National Institute of Standards and Technology).

4. Technological capabilities and optimal infrastructure of enterprises

For organizations, a technological framework is needed to allow harnessing the business capacities of big data.This not only put constraints on the existing infrastructure, but future infrastructures need agility in order to scale as needed. Here we will present our take on what such an infrastructure might consist of, and also briefly present infrastructures in a current organization, namely AT&T, as presented by John Donovan. We do note that it is not possible to give a complete overview of the inner infrastructure, but we'll aim at providing an overview of some of the pillars, however. Furthermore we'll include references on current systems within distributed computing, cloud and analytics-systems. These will be compared to article reviews and journals reflecting on the same area. We will also be drawing on Cloudera's views on enterprise infrastructure, and their presentations on low-cost frameworks.

The technological ways of tapping into big data and systematize an approach to data management and utilization will often vary from case to case, and startups and big enterprises alike has tackled this in a variety of ways. Sometimes a custom solution is needed to wholly leverage some of the potentials, exemplified by the infrastructure that AT&T has developed. As mentioned by John Donovan, AT&T had to build a grounds-up infrastructure in order to harness their data generated by users. Their data came in a massive variety, size, source, and variance. They needed a schema for how to address this manifolded issue. Not only did they need to store their data, they also needed to handle how access to it was granted, how it was categorized, how it should move, and lastly, schemas for how to interface with it. Furthermore, the sensitive data also needed to anonymized. What they ended up with, as presented by John Donovan (2014), is a framework that captures streams of data into a grid in order to systematize the data before it is put into what

he called a *"data lake"* where it can be accessed through APIs, as visualized simplistically below.

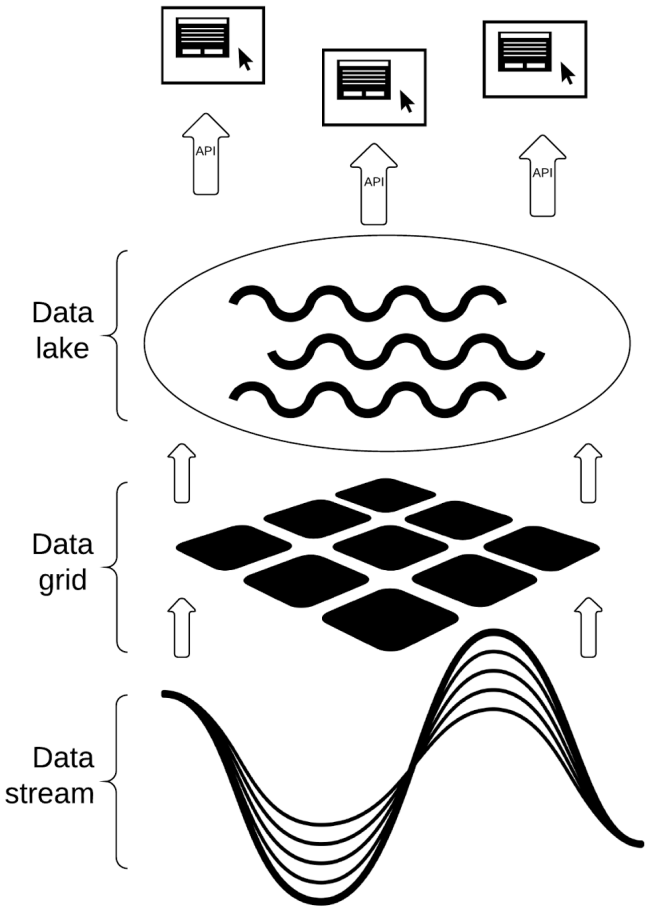


Figure 2: An overview of AT&T's data utilization schema.

But as big enterprises grasps with building solutions that covers areas of their specific use-case needs, an open source system will often be spun out, as seen with MapReduce or other systems under the Apache licence. These systems then become more easily accessible and configurable for SMEs, allowing for establishments of data infrastructures, outside that of the enterprises. It is not enough to have a lot of data; companies in general are pressed to make the rights investments on the right infrastructure, even when these have been made more accessible. This especially holds true for SMEs as their technical knowledge and economic capital can be restrictive, and the cost of investing wrongly might be fatal. As observed by Lin and Ryaboy (2011), *"a major challenge in building data analytics platforms stems from the heterogeneity of the various components that must be integrated together into production workflows…"*. A plan of action for how data should be handled and moved, under what system, must therefore be developed. The goal is to make all current and future data, in any form, available for any application in the organization, in order to reduce complexity and enhance ease of access and leverage use.

We therefore identify three distinct technical categories that forms a pivotal role in creating a low-cost data utilization framework for SMEs:

- A scalable cloud network of servers - allowing for developing a small scale operation, that can then be expanded upon.
- A framework for data management, processing and storage, such as Hadoop and MapReduce.
- Analytics software, such as Tableau, SAS, SPSS, etc.

4.1 Cloud

As SMEs may often lack the capabilities to manage and control large server and data operations themselves, we suggest a solution of public cloud servers, such as Amazon Web Services, in order to keep costs and complexity down. Cloud computing can, in short, be understood as *"... a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction"* (Mell & Grance, 2011). Cloud delivered infrastructure can help SMEs to take advantage of more sophisticated IT operations, which would not have been accessible for them due to high costs of operations and up-front investments. Where many enterprises are already bound by the limitation of existing infrastructure, opting for a Infrastructure-as-a-Service (IaaS) solution will allow SMEs to build their own platform of system management and data handling, upon the cloud delivered infrastructure. This will help them build a framework that is more agile, and with the right platform build upon, technically help them scale when needed, and allow for low costs of operations. We see that the rising importance of such a scalable cloud system might also give rise to another set of challenges, such as; data processing, data storage, and management of said systems.

4.2 Data processing and storage

The task of setting up a framework to meet and address these challenges can be a complex task for SMEs to undertake. In its simplest form there exists only a need for one database and a single data pipe. But gathering logs from sales, CRM-systems, market analysis, logistics and product performance, will require a system that encompasses a multitude of different formats and structures, to name a few, to even manage the files, however. The goal here is to allow a structure to easily manage storage and at the same time allow for a file-system to easily handle different flows and formats. Building on Cloudera's Enterprise Data Hub (Cloudera, 2014), based on Apache's Hadoop (Hadoop, 2014), we propose that this, or a similar Hadoop-based solution, is implemented as the platform for the data management, due to its track record of effective and often cheap use

in organizational setting. That is, however, only if the datasets to be handled actually is of a proportionate size and variance that such makes sense, and would otherwise have been too complex to handle with a scripting language, like Python.

Further reducing complexity of running multiple machines we propose Apache Mesos (Mesos, 2014), delivered through Mesosphere (Mesosphere, 2014), as a overlaying framework for the cloud based servers in order to run them as a single instance, distributing workloads, simplifying rolling out instances of Hadoop on select virtualized machines. Put differently, Mesos would be applied on top of the cloud-provided infrastructure of machines in order to remove multiple instances that would then need to be managed. This would allow application or database, such as Hadoop or Hive, to be run on top as a if on a single machine. This will eliminate the need for having static virtualized machines run a set number of applications, and reduce complexity for operations of databases and file-systems. Being able to run multiple applications on the infrastructure, no matter the size will help SMEs to easier and faster test different platforms and applications as needed. Most importantly, this will allow SMEs to easily scale their operations as needed, without further configurations, and to fully utilize their investment in their computing resources. The specific setup will vary based on use case. Their will however be a limit in terms of which platform and infrastructure applications that are supported by Mesos.

4.3 Analytical framework

Due to the heterogenous nature of having various components and data forms delivered for analysis the aforementioned system is critical for providing a minimum of hassle for analytical work (Lin & Ryaboy, 2011).

A question that might be posed when making sense of big data in decision making processes, is how findings can be presented in order to best convey information. As supported by SAS (2013) and Lemieux *et al* (2014) visualization of data is a key parameter that, if supported in by the right environment, can help convey meaningful insights of data faster. Reaching these insights faster are key in order to support decision making processes and stay competitive. We therefore propose analytical software with visual capabilities in order to for SMEs to focus their attention on a few but powerful analytical areas. Software such as Tableau (Tableau, 2014), IBM SPSS (IBM, 2014) or SAS (SAS, 2014), with each a very different approach, offers these capabilities. Choosing in line with Gartners magic quadrant for analytical software (Sallam et al, 2014), Tableau comes out on top with its ease of use, extensive visual capabilities, and cost below average. Allowing queries to be run across sources and types of databases eases entrance to data sources. For SMEs getting into the data analytical field, this could lower cost of operation and clearly minimize the required level of skill for managing analytical operations. For bigger enterprises however, Tableau is still mostly serving as a complimentary business intelligence tool. Gartner notes

that it also needs to address areas of business user and IT requirement for governance in enterprises (Sallam et al, 2014).

4.4 Summary

From the presented technological systems we present a merged solutions for companies to help them leverage their data. The presented was thought out as addressing the challenges that SMEs might have, but can just as well be adopted by larger enterprises. The focus on building a scalable infrastructure with maximum utilization is paramount for both. These three presented ways of approaching this forms the rationale of creating a backbone of cloud servers that allows for storage, application systems and data management tools. Secondly a system, such as the mentioned Hadoop, for managing multiple data structures and sources would be applied within the cloud service. To decrease complexity, Apache Mesos or Mesosphere could be installed as the first layer on the cloud servers to operate them as a single instance. Allowing databases, data management and applications to be installed on top, virtualizing a single pool. Providing this infrastructure of storage, data flows and management would, we argue, form a framework for applications to access multiple types of data for pre-processing of data, which in turn then allows for analytical applications such as the visual tool Tableau to be applied. As more and more companies tries to gain competitiveness through software and data-driven capabilities, starting off on the right infrastructure, build for scaling, is crucial. Both for existing enterprises and SMEs trying to move in and compete.

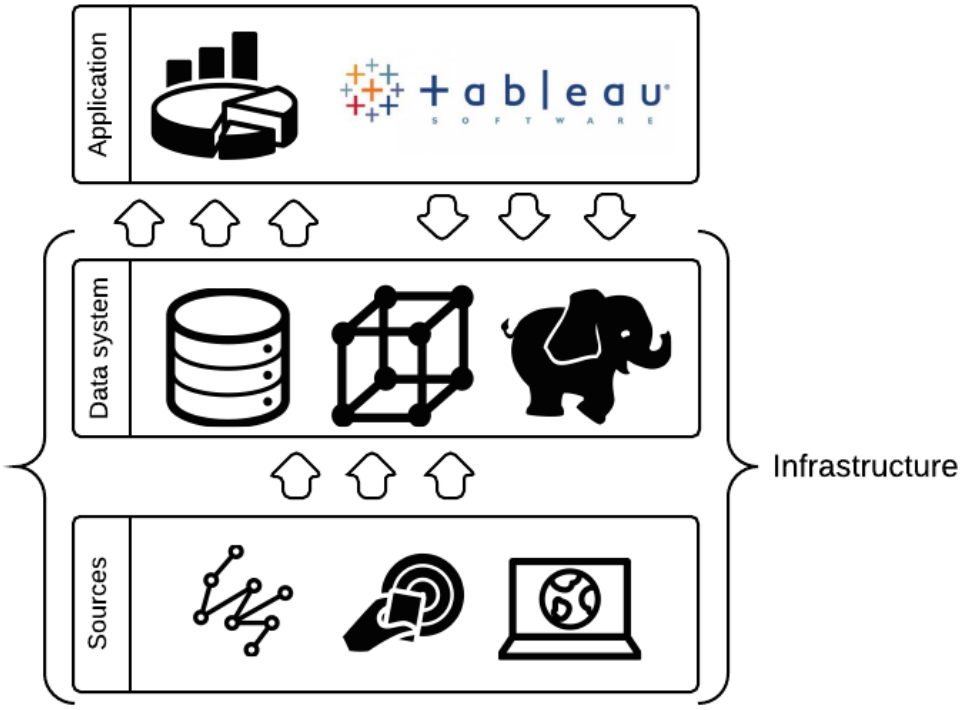


Figure 3: Our proposed framework, building upon an infrastructure gathering from data sources, which in turn is used by analytics applications.

Sources can be many and varied, but with a data system that allows for storage, effective data handling and processing through the right applications, a system can serve as a platform providing access to the right data for analytical applications when needed.

5. Organizing for Big Data on a Small Scale

"Technology does not create value; People do."

The above quotation is a common saying about technology within business. Technology is just a resource, and only when people transform resources into products or services, value is created for organizations (Christensen, 2013). Creating value from big data is no exception, and data-savvy employees are critical. However, data scientists and the like are in huge demand and in low supply (Heudecker et al., 2013). This raises the question: How can SMEs act in order to attain the necessary skills? Further, when the necessary skills are attained, how does SMEs organize to extract value? And finally, how is data or information obtained and governed on a SME scale? These are the main questions that the present part of the paper will seek to answer.

5.1 Necessary skills and profiles and how SMEs can obtain them

Big data marks the shift from stocks of fixed structured data to flow of ever growing unstructured data, and as a result, it also marks the change in labor demand from data *analyst* to data *scientists* (Davenport, Barth & Bean, 2013). There seems to be a lacking consensus on the difference, but while some argue that the difference is only in the terms used, a salary almost multiplied by 2 suggests otherwise ("Salary," n.d.). In the era of Big Data, data scientists are in high demand and low supply resulting in salaries above the 6-digit mark. With such a high competition for a skill set SMEs might ask themselves an obvious question: How should we manage to attain such skills within our organization?

To answer the question further clarification of the data scientist's skill set is needed. Davenport and Patil (Davenport & Patil, 2012) describe a data scientist as a hybrid of data hacker, analyst, communicator, and trusted advisor. In other words, the skills are from traditionally divergent areas now merged in a data focused profile. In a popular post on the question-and-answer site Quora, a data scientists from an undisclosed company working with search engine and advertising, made the following illustration to display the needed skill set for a data scientist:

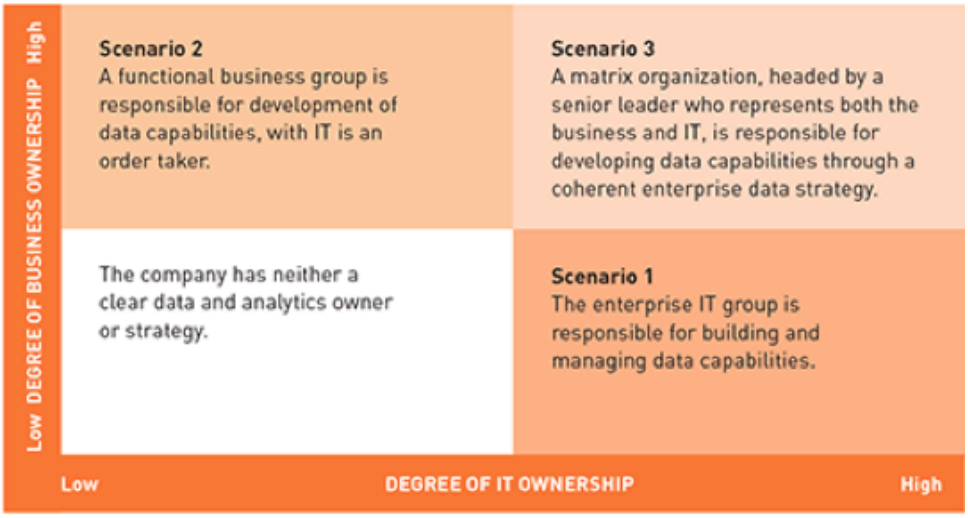
Figure 4: Skillset for Data Scientists (Iqbal, 2014)

Two aspects are important for SMEs this context. First, of all, the skill of business domain knowledge is stated as being very significant parameter in order to create value from data (Davenport & Patil, 2012), and second of all, the fact that much of the widespread and popular software used with big data is open source such Hadoop, R, Hive and D3 (King & Magoulas, 2014). Combined they set the basis for SMEs to look within the company to find potential candidates possessing some of the skills illustrated in figure X, and let them use their business knowledge to experiment with data using some of the freely available open source tools (Sicular, 2012; Davenport, Barth, & Bean, 2013; Asay, 2014;). In addition, extensive educational material for these open source tools is available through multiple Massive Open Online Courses and helpful open online communities, which allows for cheap skill development. This approach may be much more effective and value-adding than competing with the big pockets of big enterprises in the hunt for data scientist on the job market. Another approach could be to enhance business knowledge of current IT staff to leverage their logical and technological skills to work with data in the business. Further, if a SME does not possess the necessary potential in the company, they might take the approach of seeking after strong analytical skills in other fields, e.g. biology, psychology or sociology (Davenport, Barth, & Bean, 2013). All things considered, SMEs attaining the necessary skills for working with big data seems feasible if less direct methods are put into practice.

5.2 Organizing for big data

One thing is for SMEs to attain the right skill set for big data, another thing is, how SMEs might organize in order to create value from big data. And more specifically: Might it be any different from what big enterprises are doing?

A survey on big data among Fortune 500 companies was conducted in 2013, which revealed that ownership for big data initiative sometimes resided within the business side and sometimes within the technology side (Bean and Kiron, 2013). This split is also captured in the table below, created by partners of the management consulting firm formerly known as Booz & Company (Casey, Krishnamurthy & Abezgauz, 2013).



Source: Booz & Company

Figure 5: Three Data and Analytics Operating Models

Casey et al. (2013) presents three different scenarios of working with big data. Firstly, they argue that while IT departments may posses the proper technical skills, they may lack the necessary business knowledge and have a tendency to focus on the perfecting the technological solution rather than providing business value. Secondly, a business group – for instance finance or marketing - may be responsible for big data ensuring an alignment with the business strategy. However, they may not fully be able to leverage the potential in the technology, and there is the risk of silo thinking or poor architectural solutions. Finally, they suggests that business and IT can collaborate in an matrix organization headed by a competent leader with knowledge within both fields. Although this solution, Casey et al. argue (2013), posses the highest potential, it is also the most challenging from an organizational view to facilitate interdisciplinary collaboration in a more complex and expensive structural setup.

Where Leaders and Laggards Put Their Big Data Professionals		
	Leaders	Laggards
Those Who Analyze Big Data		
In IT function	42%	45%
In business functions that use the data	21%	32%
In a separate Big Data group	37%	23%
Those Who Process Big Data		
In IT function	47%	69%
In business functions that use the data	14%	11%
In a separate Big Data group	37%	19%

Figure 6: How Big Data Professionals are organized

Further, data suggests that having an actual Big Data group is pervasive among big data leaders, and evidence suggests that it provides the most benefit ("Big Data Skills- The best way to organize Big Data activities in organizations," n.d.). However, this is likely to surpass resources of typical SMEs. But what then could a solution be?

One approach could be to create small self-organized and experimenting groups with a few members covering as many of the data scientist skills previously described in a way to seek value adding benefits from big data with limited costs. Such a set-up would capture the essence of what big companies are suggested to do while not putting the viability of the company at risk

5.3 SME Governance of Information

Much of the talk around big data is centered on how the volume, velocity and variety of big data set demands for data scientist skill sets and aligned organizational structuring around a data-driven and value-adding business process. However, less prevailing seems to be the discussion of actual governing the data – or information, let alone what this means for SMEs.

Information governance is the latest term coined to capture this process. It can be defined as the "*... specification of decision rights and an accountability framework to encourage desirable behavior in the valuation, creation, storage, use, archival and deletion of information.*" (Logan, 2010). In order words, it comprises a set of rules and guidance to how data or information is handled within an organization.

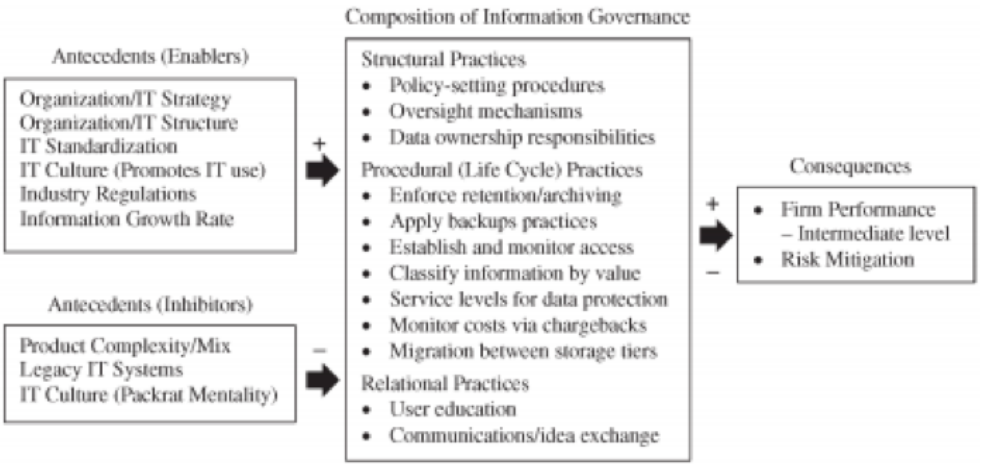


Figure 7: Overview of composition and context for the Information Governance concept (Tallon et al. 2013b)

In academic terms it is a novel concept that within the last year has been support by Tallon (2013a; Tallon et al., 2013b). Tallon et al. argues (2013b) that the extant literature and practice of IT Governance have overly focused on infrastructure, project management and control while not recognizing the value of (governing) information. With the increasing value of information in a knowledge-driven business world and focus on digital rights, data security and social responsibility, information governance seems to be a necessary part of working with big data. Especially for SMEs that might not be experienced in working with data, an information governance development seems almost obligatory. Further, Information Governance encompass the process of actual obtaining data, for instance in strategic partnerships with other SMEs (Neely, 2014), monitoring data and ensuring that value is extracted (Tallon et al., 2013b.) It may be an area that seems the least exciting from an immediate business economic viewpoint, but it might be the most important aspect to invest in for SMEs to minimize risk and ensure value if – or when - an SME decides to jump into the big data pool.

5.4 Summary

Taking big data to SMEs involves some significant challenges for organizing that may be solved with carefully executed approaches. Firstly, data scientists are scarce in a high demand job market, but by looking at potential within SMEs may leverage existing business domain knowledge and the availability to big data open source software and learning resources hereto. Secondly, SMEs may consider making a small interdisciplinary team to experiment with extracting business value from data instead of limiting it to isolated skills in the IT department or the business side. Third and finally, SMEs should strongly consider to encompass Information Governance in order to ensure data quality, security and data ownership - a formal process that might be even more important for SMEs to due to a likely lack of working with data in such a way.

6. The strategic advantages of big data

"*You can't manage what you don't measure.*"
- W. Edwards Deming and Peter Drucker

This saying has maybe never been more important in the era of big data. Managers can now measure radically more about their business and translate this enhanced knowledge into improved strategy and performance (McAfee et al., 2012). But does this apply to every market participant?

Previously, big data and the derived potential was exclusive domain of statistician, multinational enterprises and information technology departments. But the availability of data and analytics is emerging, which is called the data democratization (Intuit, 2012). This democratization enables SME access to cost efficient, useful, data driven tools and analytical systems. Intuit (2012) forecasts that the evolving data democratization will deliver meaningful insights on markets, competition and bottom-line business results for small business.

The chronology of the section is as follows. We will look into how data-driven SME's perform and how big data will underpin new waves of productivity growth. Further, the transformed approach to decision making and strategy formation will be examined. Finally the use and utilization of big data in different industries will be analyzed. First we will investigate how data has changed the approach of doing business, exemplified by Amazon.

Every company started as a startup. The Amazon story marks the power of big data and how data transformed a small startup into a massive industry disrupter and market leader. Traditional bookstores have always been able to track the books being bought. This is the data that is available to the physical retailers. Once retailing moved online, the amount of valuable data on customer buying behavior increased dramatically and created a new era of customer understanding. Amazon transformed the traditional bricks-and-mortar retailer into and a data-driven ecommerce. Amazon could not only track what customers bought, but also what they showed interest in, how they navigated the webpage, how individuals reacted to promotions and similarities across different segments. Later on Amazon developed algorithms to predict which books that are most likely for individual customers to buy next. The traditional bookstore had no chance to access this evidently valuable information (McAfee et al., 2012).

Data means nothing if it can't be understood. So not only the data collection was the key to success but the way CEO Jeff Bezos and Amazon transformed the data into actionable strategies marked the difference. The utilization of Amazon's visitor and transactional data

yielded revolutionary customer insights, which was transformed several profitable strategies such as individual target marketing campaigns (5).

6.1 Data driven decision making

Decision making and strategy formation has long been a subject of study, especially since the growth of big data. There are several definitions of Data driven decision making, but one of the more comprehensive is "*the practice of basing decisions on the analysis of data rather than purely on intuition*." (Provost, F., Fawcett, T. 2013)

A fair question to ask is: Does data driven decision making lead to improved performance and results? A study from 2011 investigated this question: "*We examine whether firms that emphasize decision making based on data and business analytics ("data driven decision making" or DDD) show higher performance*." (Brynjolfsson et al.,2011). The paper developed a measure of data driven decision making that captures business practices surrounding the gathering and analysis of internal and external data. They examined the relationship between data driven decision making and productivity and performance: *"We find that DDD is associated with a 5-6% increase in their output and productivity, beyond what can be explained by traditional inputs and IT usage."* (Brynjolfsson et al.,2011)

The results is based on publicly traded companies but due to the emerging data democratizing, we argue the data-driven decision-making approach to management is transferable to SMEs. Further, one could argue that there is a even bigger potential of performance improvement due to the assumption of more intuition based decision making within SMEs (Welter, F., 2003).

This underlines the importance and the opportunities for SMEs to invest in big data analytics to improve their foundation of decision making and strategy formation. In the following sections will we illustrate some best case examples of SMEs harnessing the potential of big data in their field of business.

6.2 Big data is changing the rules

Big data is changing the rules of customer satisfaction, commerce and business operations, generating new opportunities and challenges for SMEs. The streams of data in real time will turbocharge the use analytics in SMEs. New insights and a more profound understanding of customer needs is changing the way the general rules of doing business.

Affordable big data analytics now permits every SME to be an Amazon consumer centered company. Personalized and real time targeting will bring the personal service back. Retailing is an obvious place for data-driven customization because the volume and the

quality of data available from internet purchases, social network conversations, and more recently, location specific smartphone interactions have mushroomed (McKinsey Quarterly, 2011).

The local merchant once knew what his customers liked and preferred. But the rise of big business, mass production and hyper malls drove a majority of the small local merchants out of business. Economies of scale flourished, as did wider selection of goods and lower prices but at the cost of personalized service.

SMEs affordable access to data is now reinventing the personalized service. E-commerce SMEs can with clever use of data anticipate and meet customer needs without any human interaction. Digital footprints of customers reveal purchase patterns, preferences and interests – the same information the traditional local merchant knew through personal interactions with his customers on a frequent basis. SMEs can use this information to personalize offers and goods we want, as local businesses used to.

"Big data is bringing back personalized service by giving businesses greater insight into consumer preferences, many times without even talking to them, because customers leave digital footprints when they use the Internet, use a credit card or post on Facebook." Steve King, partner at Emergent Research (Devaney T. and Stein T., Forbes 2013)

Womply is a growing SME that strives to make it easier for local merchants to engage their best customers, personalize offerings and increase loyalty. Womply helps SMEs strengthen their relationship with their customers by combining purchase information with social media data to provide a more holistic picture of their customers.

6.3 Big data creates big opportunities

The data accessibility and improved analytics revolution are igniting opportunities for new companies and existing SMEs to find new ways to harness the power of the growing aggregation of digital data. As earlier exemplified by Amazon has the past decade of successful internet startups showed how data and insightful analytics created new unimaginable business opportunities. These immense opportunities generated by the capturing and analyzing data are creating new companies everyday.

The power of data is not limited to startups but also apply to existing business models across industries. The use of digital data and analytical tools enables SMEs to extend their products and services as well as to create new ones (Intuit, 2012).

The new wave of powerful, affordable and cloud-based tools allows novices and SMEs with a small/none analytics budget to analyze and interpret relationships though intuitively

UIs. This is one of the key drivers behind the democratization of data that it moves from the domain of advanced statisticians to SMEs and startups.

Several hotel chains have started to harness the potential of data to deliver a more personalized service to their customers. With permission from their guests they are able to recognize them through their smartphone when they are approaching the parking lot. Other applications allow guests to check in and use their smart phone as the room key. Furthermore, when customer information is aggregated, instead of being fragmented across a hotel's various divisions, the analytical insights lead to better marketing and customer service. (Hospitalitynet.org, 2013)

6.4 Summary

The emerging data democratization enables SME's access to cost efficient and useful data-driven analytics systems. This will result in new insights and improved performance for the SME's who invest time in this new field. We suggest SME's to prioritize big data from a strategically point of view based on the conclusions about better performance of data driven organizations. The next section will focus on why everything isn't a dance on roses.

7. The big data failure

This purpose of this section is to take a step back and take a more critical view on the useage of big data in SME's. There are often several reasons why the implementation of big data projects fails. We will investigate the reasons why and how SME's can mitigate the pitfalls.

Many SMEs are still in a vacuum mode where they only are exploring big data from a containment or storage scalability point of view. SMEs may be looking at smart scalable analytics, which is unconstrained in a traditional manner such as need for developer resources or databases capacity (Bodkin, R., 2013).

This means that many SMEs exploring big data are missing out on the key opportunities of big data as earlier described. To fully harness the potential and improve their business it requires investments in both people and resources.

7.1 What are the biggest mistakes?

It is no secret that the majority of big data implementations tend to fail – 55 percent of big data projects don't get completed, and many other fall short of their objectives (Infochimps, 2013).

There may be many reasons why big data projects fail. Undoubtedly one of the major factors is lack of communication between management who provide the project vision and

those who are implementing it. Traditional thinking and management impatiens is often a cocktail of failure.

Gartner's Svetlana Sicular has quantified eight causes of big data projects failures (Sicular, S. Gartner 2014). Some of the most conspicuous causes to failure are that organizations tend to ask the wrong questions. Sicular frames the problem very gently with the following statement:

"*Learning Hadoop is easier than learning the business"*
- Svetlana Sicular, Gartner

The ability to find the right insights that improves the performance of the individual business is more about the fundamental understanding of the business. Throughout Sinculars complete list one theme emerges: As much as we might want to be data driven, people is keep getting in the way. As much as we want to be directed and guided by data, people ultimately rule the data including the initial decision making of which data to collect and which questions to ask.

7.2 How to navigate the implementation?

Start small, fail fast – is a common and often successful approach, not only to big data implementation. But when it comes to big data implementation in SMEs is it the best way to go? As we've mentioned, a system might vary depending on the use case, and many organizations have specific needs. The critical point to draw from this report in terms of technical pointers, is the crucial part of building a low-cost, scalable system that allows for agile testing of systems and applications. Specifically, we have pointed towards a framework that, in order to harness analytical capabilities, can be implemented with low costs, and a low skill-gap. Including the right skill-set in implementation of such framework is still crucial, however.

As highlighted earlier, if value is to be derived from big data, asking the right questions is just as crucial as having the right tools on the right platform.

8.0 Conclusion

This project have touched upon three aspects of the challenges that SMEs, and beyond, will face when moving into the big data area. Whether it's for internal data processing, or analytical jobs. These challenges are, as shown, many-faceted, and we've tried to encompass the most central aspects within technology, organization and strategy.

Through this project we have highlighted some of the challenges that SMEs should address when utilizing Big Data. Larger enterprises face many of the same challenges, but

starting of with how setting up a framework for big data utilization in SMEs might be undertaken, we have highlighted some of the technical, organizational and strategic areas that adheres to this.

Further, we suggest SMEs to organize for big data with looking within to development the necessary skills and leverage existing business knowledge as well as the extensive free available learning resources for big data tools, to create small interdisciplinary and experimenting data teams, and to develop Information Governance to insure data quality, security and data ownership.

The emerging data democratization enables SME's to affordable and data driven analytics systems. With the right technological infrastructure and the right organizational alignment will this result in new insights and improved performance for the SME's. We strongly recommend SME's to prioritize big data from a strategically point of view based on the conclusions about improved performance of data driven organizations.

References

Agrawal, D., Das, S., & El Abbadi, A. (2011). Big Data and Cloud Computing: Current State and Future Opportunities. In Proceedings of the 14th International Conference on Extending Database Technology (pp. 530–533). New York, NY, USA: ACM. doi:10.1145/1951365.1951432

Amorosi, D. (n.d.). An inside look at AT&T's operations centre and security strategy. The Guardian. Retrieved from http://www.theguardian.com/media-network/partner-zone-infosecurity/at-and-t-operations-centre-security-strategy

Apache Hadoop. (2014). Apache Hadoop. Retrieved from http://hadoop.apache.org/

Apache Mesos. (2014). Apache Mesos. Retrieved from http://mesos.apache.org/

Asay, M. (2014). *8 Reasons Big Data Projects Fail*. *InformationWeek*. Retrieved August 15, 2014, from http://www.informationweek.com/big-data/big-data-analytics/8-reasons-big-data-projects-fail/a/d-id/1297842

AT&T and IBM Join Forces on Big Data Analytics, Cloud, Security Technologies. (n.d.). Scientific Computing. Retrieved July 27, 2014, from http://www.scientificcomputing.com/news/2014/02/t-and-ibm-join-forces-big-data-analytics-cloud-security-technologies

Bean, R. & Kiron, D. (2013). Organizational Alignment Is Key to Big Data Success. *MIT Sloan Management Review*. Retrieved from http://sloanreview.mit.edu/article/organizational-alignment-is-key-to-big-data-success/

Big Data Skills - The best way to organize Big Data activities in organizations. (n.d.). *TCS Big Data Study*. Retrieved from http://sites.tcs.com/big-data-study/big-data-skills/

Bodkin, R. (2013). The big data wild west the good the bad and the ugly from http://gigaom.com/2013/09/14/the-big-data-wild-west-the-good-the-bad-and-the-ugly/

Brynjolfsson, E., Hitt, M. L., Heekyung, K. H. (2011). Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?

Casey, T., Krishnamurthy, K., & Abezgauz, B. (2013). *Who Should Own Big Data? strategy+business*. Retrieved August 14, 2014, from http://www.strategy-business.com/article/00211?gko=44b8e

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. MIS quarterly, 36(4), 1165-1188

Christensen, C. (2013). The innovator's dilemma: when new technologies cause great firms to fail. Harvard Business Review Press.

Davenport, T. H., & Patil, D. J. (2012). Data Scientist. Harvard Business Review, 90, 70-76.

Davenport, T. H., Barth, P., & Bean, R. (2013). How "big data"is different. MIT Sloan Management Review, 54(1).

Devaney T. and Stein T., Forbes 2013 from

http://www.forbes.com/sites/capitalonespark/2013/05/30/what-can-big-data-do-for-a-small-business/
Donovan, John. (2014). Leading Trends in Information Technology, Lecture, Stanford University.
EIT ICT Labs. (n.d.). Retrieved August 14, 2014, from https://www.eitictlabs.eu/about-us/calls-tenders/call-for-smes-in-big-data-analytics-business-use-case-analysis/
Garbarino-Alberti, H. (2013). IT governance and human resources management: A framework for SMEs. International Journal of Human Capital and Information Technology Professionals (IJHCITP), 4(3), 40-57.
Gartner, (2014) Glossary. Retrieved August 14, 2014, from http://www.gartner.com/it-glossary/big-data/
Heudecker, N., Beyer, M. A., Laney, D., Cantara, M., White, A., Edjlali, R., … McIntyre, A. (2013, November 20). Predicts 2014: Big Data. Gartner Insight, Gartner Research, Stanford, Connecticut.
Hospitalitynet.org, (2013), Major global study reveals how big data will transform the hospitality industry. Retrieved from http://www.hospitalitynet.org/news/4061400.html
IBM. (2014, June 12). IBM SPSS software. CT500. Retrieved July 27, 2014, from http://www-01.ibm.com/software/analytics/spss/
Intuit 2020. (2013) The new data democracy: How big data will Revolutionize the lives of Small Businesses and Consumers
Iqbal, R. T. (2014, January 10). Data Science: What are the key skills of a Data Scientist? *Quora*. Retrieved from http://www.quora.com/Data-Science/What-are-the-key-skills-of-a-Data-Scientist
Kart, L., Heudecker, N., & Buytendijk, F. (2013, November 20). Survey Analysis: Big Data Adoption in 2013 Shows Substance Behind the Hype. Gartner Insight, Gartner Research, Stanford, Connecticut.
Kaskade J. (2013). CIO's and Big data, Infochimps Retrieved from http://blog.infochimps.com/2013/01/24/cios-big-data/
King, J. & Magoulas, R. (2014). *2013 Data Science Salary Survey*. California: O'Reilly Media.
Kooper, M. N., Maes, R., & Lindgreen, E. E. O. (2011). On the governance of information: Introducing a new concept of governance to support the management of information. International Journal of Information Management, 31(3), 195-200.
LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (n.d.). Big Data, Analytics and the Path From Insights to Value. MIT Sloan Management Review. Retrieved from http://sloanreview.mit.edu/article/big-data-analytics-and-the-path-from-insights-to-value/
Lemieux, V. L., Gormly, B., & Rowledge, L. (2014). Meeting big data challenges with visual analytics: the role of records management. *Records Management Journal*, *24*(2), 5–5.
Lin, J., & Ryaboy, D. (2013). Scaling Big Data Mining Infrastructure: The Twitter Experience. SIGKDD Explor. Newsl., 14(2), 6–19. doi:10.1145/2481244.2481247

Logan, D. (2010, January 11). What is Information Governance? And Why is it So Hard? Retrieved from http://blogs.gartner.com/debra_logan/2010/01/11/what-is-information-governance-and-why-is-it-so-hard/

Massachusetts Institute of Technology. (2014, January 30). Storage system for "big data" dramatically speeds access to information. ScienceDaily. Retrieved July 27, 2014, from http://www.sciencedaily.com/releases/2014/01/140130111001.htm

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big Data. The management revolution. Harvard Bus Rev, 90(10), 61-67.

McKinsey Quarterly, (2011). Are you ready for the era of 'big data'?

Mell, P., & Grance, T. (2011). The NIST definition of cloud computing.

Mesosphere. (n.d.). Retrieved from http://mesosphere.io/

Neely, A. (2014, April 20). Even SMEs need big data. Retrieved from http://www.telegraph.co.uk/technology/technology-topics/10772462/Even-SMEs-need-big-data.html

Oz, E. (2008). *Management information systems*. Cengage Learning.

Pearlson, K.E and Saunders, C.S. (2006) The Information Systems Strategy Triangle. I Managing & Using Information Systems - A Strategic Approach (s. 19-37). Hoboken, NJ: Wiley.

Provost, F., Fawcett, T. (2013) Data Science and its Relationship to Big Data and Data-Driven Decision Making.

Revolution Analytics. (2011). Advanced "Big Data" Analytics with R and Hadoop (White Paper). Revolution Analytics. Retrieved from http://www.revolutionanalytics.com/sites/default/files/r-and-hadoop-big-data-analytics.pdf

Salary: Data Scientist. (n.d.). Glassdoor. Retrieved August 15, 2014, from http://www.glassdoor.com/Salaries/data-scientist-salary-SRCH_KO0,14.htm

Sallam, R. L., Tapadinhas, J., Parenteau, J., Yuen, D., & Hostmann, B. (2014, February 20). Magic Quadrant for Business Intelligence and Analytics Platforms. Gartner. Retrieved from http://www.gartner.com/technology/reprints.do?id=1-1QLGACN&ct=140210&st=sb

SAS. (2013). Five big data challenges And how to overcome them with visual analytics. SAS. Retrieved from http://www.sas.com/resources/asset/five-big-data-challenges-article.pdf

SAS. (2014). *SAS Products - Business Intelligence and Analytics*. Retrieved from http://www.sas.com/en_us/software/business-intelligence.html

Sicular, S. (2012). *Data Scientist – Mystified*. Retrieved from http://blogs.gartner.com/svetlana-sicular/data-scientist-mystified/

Sicular, S., (2014). Big botched data, Gartner Institute. Retrieved from http://blogs.gartner.com/svetlana-sicular/big-botched-data/

Staten, J., (n.d.). AT&T turns its data into an adaptive intelligence business. ZDNet. Article. Retrieved July 27, 2014, from http://www.zdnet.com/at-and-t-turns-its-data-into-an-adaptive-intelligence-business-7000021348

Tableau. (n.d.). *Tableau Software*. *Tableau Software*. Retrieved August 6, 2014, from http://www.tableausoftware.com/

Tallon, P. P. (2013a). Corporate governance of big data: Perspectives on value, risk, and cost. *Computer*, *46*(6), 32-38.

Tallon, P. P., Ramirez, R. V., & Short, J. E. (2013b). The Information Artifact in IT Governance: Toward a Theory of Information Governance. *Journal of Management Information Systems*, *30*(3), 141-178.

Welter, F., (2003). Strategy, SME & environment - patterns of action and strategy in SMEs. Berlin: Duncker & Humblot.