How can SMEs benefit from Big Data? Challenges and a path forward

Shirley Coleman, Rainer Göb, Giuseppe Manco, Antonio Pievatolo, Xavier Tort-Martorell and Marco Seabra Reis
(Short Bios at the end of the paper)

Abstract

Big data is big news and large companies in all sectors are making significant advances in their customer relations, product selection and development, and consequent profitability through using this valuable commodity. SMEs have proved themselves to be slow adopters of the new technology of big data analytics and are in danger of being left behind. In Europe, SMEs are a vital part of the economy and the challenges they encounter need to be addressed as a matter of urgency. This paper identifies barriers to SME uptake of big data analytics and recognises their complex challenge to all stakeholders, including national and international policy makers, IT, business management and data science communities.

The paper proposes a big data maturity model for SMEs as a first step towards an SME roadmap to data analytics. It considers the "state-of-the-art" of IT with respect to usability and usefulness for SMEs and discusses how SMEs can overcome the barriers preventing them from adopting existing solutions. The paper then considers management perspectives and the role of maturity models in enhancing and structuring the adoption of data analytics in an organisation. The history of Total Quality Management is reviewed to inform the core aspects of implanting a new paradigm. The paper concludes with recommendations to help SMEs develop their big data capability and enable them to continue as the engines of European industrial and business success.

Keywords

predictive analytics, maturity model, data science, skills shortage

1 INTRODUCTION

An enormous amount of data is created worldwide every day. Much of this data is directly or indirectly relevant for policy and decision making in: industrial enterprises, especially in supply chains, manufacturing and service processes; company and customer management; financial and commodity markets, and macroeconomic development. Thus most business activity now takes place in the "big data" environment and indeed "big data" is essential for the survival of many companies. "Big data" has four dimensions, often addressed by the acronym VVVC[1]: high-volume (V) data, created with high velocity (V) in great variety (V) and of high complexity (C). Regarding volume, 2.5×10^6 terabyte of data are created worldwide per day. It is expected that the creation volume is doubling every 40 months. Regarding velocity, modern IT infrastructures enable data to be submitted for analysis in nearly real time. Regarding variety, the various data sources bring about data of different formats, like classical database formats, textual data, image data, sensor data, semi-structured and totally unstructured data. Regarding complexity, important aspects of data complexity are multivariate, multiformat, multirate, and multiresolution. A large number of covariate data for one target entity, e. g., a sales process, arise from different sources in different formats at different rates of acquisition and granularity.

For a long time, industry and business have just utilized data from the point of view of recording and monitoring transactions. A broader awareness of the analytical value of data has been gaining ground in recent years, and is reflected by the growing interest in business analytics. The term "business analytics" denotes the totality of data-based inference methodology used for the objective of analysing, predicting, and controlling processes in business and industry. Three subcategories of business analytics are distinguished: i) Descriptive analytics summarise, condense, and aggregate data in a way to make big and complex data sets more easily accessible for human

understanding. Graphics and aggregate statistical metrics are the basic tools of descriptive analytics. ii) Predictive analytics enable forecasts of future effects based on historical data. The methodology of predictive analytics comprises statistical learning, machine learning (ML), data mining and knowledge discovery from databases (KDD). iii) Prescriptive analytics transforms the results of descriptive analytics and predictive analytics into business decisions. Methods from optimisation theory and operations research are important quantitative tools of prescriptive analytics. Work in these areas is referred to as data science with people proficient in data science being referred to as data scientists.

The business analytics consulting services and software market is growing rapidly. Global organisational spending on analysing big data is expected to grow at rates of nearly 30 % per year until 2018[2]. Recent studies expect an annual growth rate of around 10 % for the business analytics software market[3,4].

The level of adoption and refinement of data analytics, however, varies considerably over industrial sectors. A recent study by KPMG[5] sees the insurance industry as the leader in the usage of advanced data analytics and big data strategy. Two of the earliest pioneers of harnessing big data were supermarkets and banks. Supermarkets used big data to determine what products people bought together, to identify fluctuations in sales of items in certain days, weeks or months, and to understand the sort of offers or advertising to which people react positively. Many banks have been using big data to rate the creditworthiness and predict the solvency of clients. Financial market participants use big data analytics in High-Frequency Trading (HFT) algorithms to make automatic transaction decisions[6]. In this regard, HFT firms are believed to account for more than 70% of all trading volume in US equities, 40% of volumes in US futures and 20% of volumes in US options[7].

A high level of refinement of data analytics is also present in the energy and retail sectors. In some sectors, advanced data analytics are concentrated in specific company functions, e. g. supply chain in the automotive industry. Underperformers are the media sector and the machinery and equipment sectors.

SMEs are lagging behind in the usage of business and big data analytics. In 2012, the adoption rate of big data analytics among UK SMEs was only 0.2 per cent, compared to 25 per cent for businesses with over 1,000 employees[8]. Market studies expect an annual growth rate of the global SME big data market by 42 percent over the period of 2013 until 2018[9]. However, since they are starting from a drastically low level, big data adoption in SMEs will continue to lag behind the evolution in large companies.

This paper first considers reasons for the poor adoption of big data analytics by SMEs and notes the importance of SMEs in the European business community. It then proceeds to consider IT and management issues and constructs a table from which a bespoke SME big data analytics maturity model is proposed. Actions that follow from the diagnosis using the maturity model are then discussed in terms of company organisation, deployment of data scientists, acquisition of expert support including training, and considerations of hardware architecture requirements and software solutions. The paper concludes with recommendations aimed at a disruptive change in big data analytics uptake by SMEs in Europe.

2 PROBLEMS AND CHALLENGES IN BUSINESS ANALYTICS AND BIG DATA ANALYTICS FOR SMEs

Various factors condition the poor adoption of business and big data analytics by SMEs. Among these, we have identified the following ones, as being more pervasive and relevant.

i) **Lack of understanding**: The e-skills UK[8] survey highlights an extremely low understanding of big data analytics by SME representatives, whereas among the

representatives of larger organisations around 30 to 40 % claim to have good or very good understanding of big data analytics. It is clear that SMEs will not step into a domain which they seemingly do not understand. A similar result was obtained by a survey among German SMEs where only around 30 % of the respondents considered their knowledge of big data to be good or very good[10]. It is clear that SMEs will not embark on a paradigm which they are not sure of. Most SMEs are unsure whether their data has at least one of the big data dimensions, and therefore whether investing in data science is going to bring the benefits claimed by big data enthusiasts.

ii) **Dominance of domain specialists**. Operating in a highly specialised field is a particular strength of many SMEs. The major part of the staff are domain specialists. General management functions are poorly covered. Hence there is reduced awareness of new business trends and opportunities, such as business and big data analytics.

iii) **Cultural barriers and intrinsic conservatism**: Domain specialised SMEs often used to have little interest and confidence in management trends. This attitude can lead them to classify business and big data analytics as a management hype rather than as a perspective opportunity. Another issue is infrastructure, few organisations set themselves up purely as data industries, thus when they decide to commit to a big data project, they realise that their data is not accessible or in the format that is desired; when a data specialist comes into the organisation, it is not a simple task to extract value from the data.

iv) **Shortage of in-house data analytic expertise**: Most SMEs have few or no in-house data-analytic expertise to approach advanced big data analytics. Various factors hamper the creation of adequate in-house expertise. a) High set-up costs relative to uncertainty in future returns from data analytics. b) Lack of management expertise to design, establish, and monitor a data analytic unit. c) Shortage of qualified workers, excessive staff costs, see v). In a survey among Austrian SMEs, the lack of in-house data analytic expertise was identified as a major barrier against the take-up of big data projects[11]. The hampering effect of lacking in-house expertise is very similar with

respect to the adoption of data analytics and with respect to IT adoption. In the latter respect, numerous studies are reviewed by Ghobakhloo et al.[12].

v) **Bottlenecks in the labour market**: There is a growing shortage of qualified data analysts on the labour market. Manyika et al.[13] predict that in 2018 "the United States alone faces a shortage of 140,000 to 190,000 people with analytical expertise and 1.5 million managers and analysts with the skills to understand and make decisions based on the analysis of big data". The e-skills UK[8] study expects that the demand for big data specialists will increase from 2013 until 2018 by 243 %. Even large companies have difficulties filling the intended positions. The e-skills UK[8] survey reveals that 57 % of all recruiters experienced difficulties in filling big data analysis positions. The dearth of data scientists is hindering the further development of the European market of predictive analytics services, see Probst et al.[14]. Salaries have been increasing to levels which are not accessible for most SMEs. The salary information service payscale[15,16] reports an average salary of $93,883 per year for data scientists, versus an average salary of $71,699 per year for software engineers. The requested profile of skills required tends to be more sophisticated for SME posts than at large companies; large companies can afford to distribute functions over several persons. An SME would look for cross-sectional expertise in business, IT, computing, and data analytics. Such a profile is even harder to find on the market.

vi) **Lack of business cases**: The availability of exemplary case studies and success stories is an important factor for the successful propagation of innovation in business and industry. Although guidelines and examples exist, for example in Ahlemeyer-Stubbe et al.[17], stimulating and trend-setting big data SME usage cases are not widely available. In a survey on German SMEs, a large majority of respondents emphasised the need for big data use cases, stating at the same time that existing use cases, such as those generated in the EU funded BYTE project, often do not correspond to the interests of SMEs[10,18].

vii) **Shortage of useful and affordable consulting and business analytics services**: A major part of consulting services used by SMEs concerns the operational level, e.g.,

accounting or hardware and software related IT issues. Management and business analytic consulting is less considered by SMEs. One major reason is that the management consulting sector is dominated by large consulting companies whose business practices are not in line with SMEs' needs and financial capacities. Felber[19] describes the goal of larger consulting companies as selling teams of consultants to undertake complex projects for extended periods of time. Such service is unaffordable for many SMEs.

viii) **Non-transparent software market**: Plenty of business analytics software solutions exist on the market. For users with little or no expertise it is hard to select a product with a good price-performance ratio and to separate the wheat from the chaff. The existing comparison and evaluation platforms are strongly vendor-biased. Independent evaluations and selection schemes are hard to find.

ix) **Lack of intuitive software**: The present market offer in business and big data analytics is split into two extreme parts: potentially useful but highly complex solutions requiring the expertise of knowledgeable data scientists, and some simple but less effective implementations. Solutions with both an intuitive user interface and a strong analytical potential are rare. IBM's Watson Analytics is one of the few exceptions. Market analysts emphasise the need for predictive analytics software with intuitive user interfaces and a shorter learning curve[14].

x) **Lack of management and organisational models**: To make business analytics an economic success, a company needs an appropriate management concept and organisational structure. Management challenges in business analytics and big data analytics have been addressed in the literature[20]. Organisational issues have been considered particularly in the context of maturity models[21]. However, the discussion hitherto concentrates exclusively on the requirements of large companies. For instance, issues like leadership, allocation to departments, horizontal and vertical relationships, centralised versus distributed functions have little or no relevance for SMEs. The suggested maturity models have rather an assessment purpose than the purpose of

providing constructive detailed advice on how to build up and maintain business analytics in a company.

xi) **Concerns on data security**: Data security concerns are a key obstacle in the SMEs' path to big data analytics. In an international survey among 82 companies[22], about 50% of the respondents identified data protection and data security concerns as a barrier for big data analytics. The data security issue is more serious for SMEs than for larger companies. In general, the conditions of and the expertise in IT security are at a lower level in SMEs than in bigger companies[23]. An important security gap at SMEs is the use of outdated and unsupported database management systems. Microsoft Windows Server 2003, a major platform used by of SMEs, is a notorious example of the aforementioned situation since Microsoft is ending regular support for that software in mid-2015[24]. Consequently, SMEs are more exposed to data breaches and are more vulnerable to intrusion and cyber-attacks. According to recent surveys, 80% of SMEs cyber-attacks resulted in PCI (Payment card industry) compliance fines, 62% of breaches were targeted at SMEs, 60% of these close within six months of an attack and 40% of all targeted cyber-attacks were directed to SMEs[25,26]. The big data environment implies further challenges. Large volumes of data are transmitted through multi-user and multi-owner channels, particularly in supply chains. Being unable to create an in-house data analytic environment, SMEs will resort to outsourcing analytic services, with a further loss of control over data. Security concerns are particularly serious with respect to cloud services.

xii) **Concerns on data protection and data privacy**: Customer data processing and analysis has to obey legal constraints on data protection and privacy. In 2012, the European Commission initiated a comprehensive reform of the data protection rules in the EU which should lead to a single law, the General Data Protection Regulation (GDPR). The reform should be completed by 2015. The present EU data protection regulations and their implications are considerably intricate and not easily accessible for judicially untrained persons. The Handbook on European data protection law[27] has over 200 pages. Many SMEs cannot afford the expert lawyer support needed to understand all the requirements of the legislation.

xiii) **Different venture concept:** The business model for SMEs is often built around specific market opportunities or the existence of differentiating skills and strategic

resources that makes them competitive in the local or global market. This focused venture perspective creates the idea that business is only dependent on the way they excel in such dimensions, eventually overlooking other resources at their disposal, as well as new opportunities to improve and diversify their activity.

xiv) **Financial barriers**: Numerous studies have identified financial barriers as a major obstacle for SME growth[28,29,30,31,32] for instance. SMEs have less access to debt finance than larger companies, particularly due to imperfect or asymmetric information between financial institutions and SMEs[33]. Limited financial resources cause SMEs to be very cautious about new investments beyond their specific business scope. Studies on the hampering effect of limited financial resources on IT adoption in SMEs are reviewed by Ghobakhloo et al.[12].

Delayed innovation and incapacity to adopt new technologies in the SME sector are critical for the European economy because of the large representation of the SME sector: In 2013, the EU28 nonfinancial business sector counted for 21.6 million SMEs with 88.8 million employees and €3,666 trillion in value added. This means that 99 out of every 100 businesses are SMEs, as are 2 in every 3 employees and 58 cents in every euro of value added[34]. This structure is perfectly visible in the German economy: In 2012, 99.3 % of companies were SMEs and 60 % of all employees worked in SMEs[35].

The European Commission[36] has realised the need to implant data analytics in the SME sector and stated:

"But SMEs themselves also need to embark on a cultural change if they are to exploit the potential of Big Data. This requires them to investigate data-handling tools and methods outside their small structures, and be prepared to use Big Data actively in their decision-making processes. They need to be ready to dive in and explore the growing ocean of information that is waiting for them out there."

However, exhortations alone won't help to put SMEs on the way to data analytics. The above identified problems imply a complex challenge to all stakeholders, namely national and international policy makers, the IT community, the business management community, and the data science community. The subsequent sections elaborate on some core aspects of an SME roadmap to data analytics. Section 3 screens the state of the art of IT under the perspective of usability and usefulness for SMEs. Section 4 discusses ways and means of overcoming the barriers which are preventing SMEs adopting the existing IT solutions. Section 5 considers the path forward: in terms of the role of maturity models in enhancing and structuring the adoption of data analytics in an organisation and the organisational models for business and big data analytics in an enterprise; and section 6 identifies core aspects of the process of implanting a new paradigm in business management by reviewing the history of Total Quality Management. Section 7 summarises the outlook and gives some conclusions and recommendations.

3 IT: STATE OF THE ART

A study published in MIT Sloan Management Review[37] affirms the existence of a "strong correlation between driving competitive advantage and innovation with analytics and a company's effectiveness at managing the information transformation cycle, that is: capturing data, analysing information, aggregating and integrating data, using insights to guide future strategy and disseminating information and insights". This approach represents an additional challenge to those described above to be embraced by SMEs. In particular, the process of producing valuable information for end users requires infrastructures for gathering, processing and managing huge volumes of data. There are four essential challenges, as pointed out by Labrinidis et al.[38], which should be handled in the data management life-cycle: Acquisition, Organisation, Analysis and Decision.

Acquisition. Big Data architectures are required to acquire high-velocity data from a variety of sources and to deal with diverse access protocols. Much of this data should also be filtered or compressed without compromising the ability to reason about the underlying activity.

Organisation. The acquired data may come in heterogeneous formats which are not always ready for the analysis. Hence, big data architectures require additional tools to organise data in structured formats where relevant information (such as entities, relations among them, etc.) is extracted, transformed and integrated. . These tools help the analyst in resolving (semantic) heterogeneities issues.

Analysis. This challenge concerns the modelling of the data through querying mechanisms as well as the exploitation of ad-hoc algorithms to find new insights. Methods for querying and mining big data are fundamentally different from traditional statistical analysis on small samples, as they need to be able to deal with noisy, dynamic, heterogeneous, untrustworthy data and data characterized by complex relations. In addition, supporting query processing and data analysis requires scalable mining algorithms and powerful computing infrastructures

Decision. Being able to take valuable decisions is a consequence of the efficient interpretation of results deriving from the analysis of big data. This may require users to be able to analyse the assumptions at each step of the data management process and their effects on the results.

In order to support the above challenges, there is a standard reference architecture for big data analytics, which is characterized by a multi-layered structure as depicted in figure 1.

Figure 1 here

Figure 1: Reference IT infrastructure for Big Data Analytics

The architecture exhibits 7 main layers.

i) **Hardware and networking infrastructure**, consisting of the physical connected devices providing the core computational power and memory requirements.

ii) **Data storage,** representing the software components for the storage and management of large data repositories. We can further distinguish two different sublayers:

- **File Systems**- Distributed file systems providing storage, fault tolerance, scalability, reliability, and availability. Core technologies in this layer are for example Google File System (GFS) and Hadoop Distributed File System (HDFS)[39,40]

- **Data Stores**- These are the evolution of traditional application databases, which guarantee high performance distributed access and querying of data into heterogeneous formats. NoSQL databases feature flexible modes, support for simple and easy copy, simple application programming interfaces (APIs), eventual consistency, and support of large volumes of data. NoSQL databases are becoming the core technology for data stores[39]. Common NoSQL categories are: Key-Value databases, such as Cassandra or Dynamo[41,42]: Document databases, such as MongoDB or CouchDB[43,44]; Column-oriented database, such as HBase and BigTable[45]; and Graph databases, such as Neo4J[39].

iii) **Data Processing**, representing the core technologies and frameworks for Streaming, Interactive, Real Time, Batch and Iterative data processing. These frameworks work on top of the data storage infrastructure, and provide support to more complex data analytics and integration components, by enabling scalable primitives for the access and management of data. A paradigmatic example in this layer is the widely adopted batch computational framework *MapReduce* [46], a simple but powerful programming model for large-scale computing through automatic parallel processing and distribution. In MapReduce, the computing model only has two functions, i.e., *Map* and *Reduce*, both of which are programmed by users. The Map function processes input tuples and generates intermediate key-value pairs. Then, MapReduce will combine all the

intermediate values related to the same key and transmit them to the Reduce function, which further compresses the value set into a smaller set by processing value lists.

iv) **Data Integration**, representing the supporting technologies for data ingestion, extraction, transformation, loading (ETL) and metadata management.

v) **Data Analytics**, i.e. analytical tools and libraries, which support exploratory, descriptive, predictive, statistical analysis and machine learning. The analysis tools range from declarative languages like SQL to procedural languages like Pig. Besides that, this layer includes libraries and systems supporting out-of-the-box implementations of the most common data mining and machine learning libraries. (such as R, Matlab, Mahout, MLLib).

vi) **Visualisation and reporting**, for outputting results in support of the interpretation phase.

vii) Finally, the orthogonal layer of **Infrastructure management** the operational frameworks for security, benchmarking and performance optimisation to manage workloads, resource scheduling and management and activity coordination.

The complexity of the above architecture clearly represents an apparent barrier for SMEs, both from a financial and cultural point of view. Nevertheless, there are two main factors which can provide a viable bridge between SMEs and big data analytics. The evolution of cloud architectures, as well as the development of open source Big Data Analytics projects, can help SMEs to gain insights about the related core technologies and tools.

Table 1 here

According to Mell and Grance[47], cloud computing is defined as a model for on-demand network access to a shared pool of configurable computing resources. The solutions offered can be be classified into three main categories[48], depending on whether they offer specific software services, platforms or core infrastructures. Table 1 provides a summary description of the core functionalities offered by these categories.

SMEs can take a lot of benefits from embracing business cloud solutions, the most important being access to extensive environments that potentially give enough leverage

to compete with the bigger organizations in the industry. In fact, cloud solutions propose a flexible model which ensures: platform-independence, portability and flexibility; simplified management and control through instantaneous switching of system configurations; automatic software upgrades and license management, which ensures the latest version, features and security patches, as well as high levels of reliability and protection through customisable security protocols. Further, from an infrastructural point of view, they enable access to unlimited CPU and memory resources, which can be expanded, scaled and reduced anytime, for cost effettiveness. This gives the possibility to setup hundreds of machines for creating personalised clusters, parallelising massive jobs, and then shutting it down once done.

Cloud platforms represent a cost-effective way of building the hardware and data infrastructure. The current market offers several solutions in this respect[48]. Some paradigmatic examples are Amazon Web Services (AWS), Microsoft Azure, and Google Cloud DataLab. All of these vendors qualify as technology leaders in Gartner Magic Quadrant on Cloud Infrastructure as a Service, and provide affordable and powerful solutions on a pay-per-use basis. Interestingly, the cost model they propose is based on Pay-per-usage, a business model that offers flexible prices and allows organizations to pay only for the consumed cloud resources which may vary according by level of demand. For example, a configuration including a set of instances capable of running data analytic applications ranges from from $0.336/hour to $7/hour on Amazon.

Besides the availability of cloud infrastructure solutions, it is interesting to notice that many of the best known big data tools available are open source projects. Well-established open source solutions exist for each of the technological layers mentioned above. As noted by Harvey[49],*"The very best known of these is Hadoop, which is spawning an entire industry of related services and products. The Apache distributed data processing software is so pervasive that often the terms "Hadoop" and "big data" are used synonymously"*.

Figure 2 in here

Figure 2: An example of verticalisation based on the Amazon EC2 infrastructure and open-source components.

Figure 2 shows an example of a platform, built on top of an Amazon Cloud infrastructure, and relying on Open Source/Apache licensed software. All of the architectural layers here are instantiated with Open-Source solutions which can work cooperatively. This example clearly represents a big opportunity for SMEs which can hence envisage platforms that would easily fit their existing budgets.

4 IT: PROBLEMS AND CHALLENGES

Section 3 demonstrates that affordable hardware and software solutions exist. This being the case, section 4 asks why SMEs don't embark on these opportunities.

Indeed, the abundance of affordable hardware and software solutions particularly tailored for SMEs does not represent an opportunity by itself. There are some crucial issues that need to be investigated and solved. Providing solutions to such issues can remove the technological barriers and leverage the adoption of big data analytics in the SME context.

Data security is a major issue which is particularly related to the adoption of cloud infrastructures. Proprietary data transferred to external storage systems can be a concern, especially when the data contain sensitive information. An in-house solution, based on building their own infrastructure for storing and manipulating data clearly does not suffer from this issue, but again one has to weigh the costs for setup and maintenance. In this respect, it is necessary to investigate guidelines and practices that can help a SME to make the correct decision based on affordability, privacy and

security concerns. From a technical viewpoint, hybrid solutions and practices based, e.g. on encryption or anonymization techniques could be envisaged and provided as part of the analytics expertise and organisational models.

The matching of **functional requirements** and **solutions** is a main issue. As mentioned by Loshin[50], it is challenging to differentiate Big Data Analytics products based on functionality alone, as many of the tools share similar features and capabilities and several tools or infrastructures exhibit extremely subtle differences. It is necessary to be able to correlate ease of use and algorithmic sophistication to the internal organisation's capability and level of maturity in analytics. In particular, the choice for the right big data analytics infrastructure should rely on factors such as

- Analyst expertise and skills.
- Analytical diversity.
- Scope of the data to be analysed.
- The quest for collaboration and product integration.

Also, **Usability** represents a big impedance issue. If we consider the intrinsic conservatism which characterises the cultural environment of SMEs, it is clear that the learning curve and the ease of use of software tools as well as technological infrastructure should be carefully evaluated. Particular regard should be given to software solutions which do not require special skills in order to put them to work. Another critical issue is interactiveness, i.e., the capability of allowing prompt and adequate user interaction such as feedback/interference/guidance from users. Higher interactiveness boosts the acceptance of a complicated analytics system and its mining results by potential users with little or no background.

The **development of appropriate skills** is also problematic in this setting. The complexity of the IT infrastructure as depicted above, requires a set of expertise which spans from system setup and configuration to database administration and management, programming knowledge (especially in distributed and parallel environments) and research, analytical, interpretative and creative skills. Within large organisations and projects, these skills are distributed among several individuals in large teams. By contrast, SMEs cannot afford large teams, and hence these skills should be guaranteed

by (small) teams where individuals cover several aspects in conjunction.

There is also a challenge relative to the **availability of data**. The effectiveness of the technological infrastructure depends on the actual capability of ingesting big data into the system. Notably, data of interest can be *human-sourced* (representing records of human experience such as, e.g., social network data), *process-mediated* (coming from traditional business systems, such as e.g. medical records or commercial transactions) or *machine-generated* (representing measures and events from the physical world, such as e.g. mobile tracking, weather, security logs). Often the development of a big data analytics application requires the combination of different types of data, which are not necessarily owned by the companies. Nowadays data is available from a wide variety of sources spanning from online databases of public statistics, to data coming from marketplaces (such as Microsoft's Azure Marketplace or Infochimps). However, availability does not necessarily pair with usability of such data, as the latter often comes in noisy or difficult to interpret formats. Even the internal sources can be difficult to manage when they are heterogeneous or machine-generated. It is clear that SMEs need guidance on how to manage such data. This is essentially a challenge, which includes a combination of technical skills, knowledge of specific business cases and specific organisational models.

Tuning Cost and Performance is still an open issue, even for big enterprises. We've seen that there is a plethora of architectural choices, which reflect on the performance of the overall infrastructure and likely affect the setup and maintenance costs. Unfortunately, the choice of the most appropriate configuration is still an open issue, which is approached empirically. There is no support for estimating the cost or duration of a computation before performing it. Users must do their own measurements and adjustments on the run, and tune the architecture accordingly. In short, it is still difficult to estimate the effects of different potential configurations on task and computation time. This difficulty can be fatal for SMEs, which work with very limited budgets and expertise.

5 PATH FORWARD

Before engaging in a companywide big data programme, the current state of maturity regarding the proper management and use of data should be first assessed. Also, some strategic decisions need to be made at an earlier stage, to operationalize the initiative and define the organisational model to be implemented. These topics will be addressed in this section.

5.1 ASSESSING THE COMPANY'S MATURITY

The effective implementation of big data programmes in organisations requires a preliminary assessment of their maturity regarding the strategic use of data in the daily processes. The pre-assessment of the organisation's readiness and maturity towards the use of data can be made in a systematic way by adopting a proper reference system that contemplates a gradient of well-characterised scenarios ranging from ad hoc practices to highly structured and optimised processes. Using such reference system, also called maturity model, it is possible to identify and diagnose the current state of affairs regarding data exploitation in an organisation, from which a development and implementation plan for big data can be properly designed. This activity is reminiscent of consultancy companies operating in the scope of implementing company-wide big data programmes, and therefore it is not surprising that most of the currently known maturity models for big data have been put forward precisely by representatives of this sector. Analysing the variety of maturity models proposed by a rich variety of active players in the big data movement (IBM, SAP, IDC, HIMSS, TDWI, etc.), one can verify that they base their assessment of the organisation's maturity regarding the adoption of big data according to a subset of the following dimensions:

i) **Business strategy**: the level at which business strategy is taken into account in the development of data and analytics infrastructure, for instance to enrich customer engagement, improve operational efficiency, innovate or tune business models.

ii) **Data management**: how efficient is the process of data collection, storage and retrieval managed, not only regarding transactional data from inner sources (well-structured records) but also from market and other external sources (unstructured data).

iii) **Existence of specialised people and analytical skills**: the level of skills and know how in the company, regarding the different aspects of implementing effective big data programmes (e.g., database management, hardware, analytics, etc.).

iv) **Technological infrastructure**: the installed hardware architecture that provides the company with the capability to collect, store and transmit data.

v) **Level of enterprise adoption**: how extensive is the company engagement in data-centric management (from localised efforts to company-wide implementations).

vi) **Leadership and corporate culture**: how supportive and motivating is the company leadership and established culture, towards the effective use of data for running the operations and business processes.

vii) **Data governance**: relative to the existence of effective policies for data distribution and usage (including security, privacy and ethical issues), lifecycle management, and for ensuring the access to high quality information.

The maturity level of a company depends on its ability to effectively address and overcome the problems and challenges referred to in Section 2, and therefore they should be somehow reflected in the aforementioned 7 dimensions that characterise a maturity model. Table 2 points out the dimensions of the maturity model that are particularly related to each challenge, or that are impacted by it.

Table 2 here

Based on the set of independent assessment dimensions presented above (inputs), a hierarchy of maturity levels can be proposed (outputs). Most often, maturity models contain five categories. These mappings (the core of the maturity models) are usually proprietary frameworks from the organisations that develop them, and their inner structure is not made available to the general public. On the other hand, there is no

general agreement in the typology and exact designation of the levels used. In fact, we have verified that different proposals of maturity models address distinct perspectives regarding the adoption of big data, for which the organisation maturity is to be assessed. In particular, our analysis identified the following distinct perspectives and associated levels:

i) **The process perspective**: focused on the processes implemented for managing all aspects of data usage. Levels: Ad hoc, Localised, Repeatable, Managed, Innovative.

ii) **The functional perspective**: for assessing the impact of data-centric activities across functional areas with increasing scope in the company's activity. Levels: Performance management, Functional area excellence, Value proposition enhancement, Business model transformation.

iii) **The corporate adoption perspective**: the extent of the company's engagement in the use of all available data sources for conducting its activity. Levels: Pre-adoption, Early adoption, Corporate adoption, Mature/visionary adoption.

All the models referred to above were developed and calibrated for large enterprises. However SMEs operate in very different contexts of internal resources available (financial, human and technological) and market targets, and therefore the maturity models should be developed, tuned and recalibrated to better reflect their reality, namely the importance of the changing agents, the key role of leadership, a healthy relationship with external service providers (as cost-effective solutions for the otherwise high investments needed to acquire technological and data management infrastructures), among others. A new maturity model for SMEs is therefore needed, in order to provide a consistent and effective way to define the organisation's readiness and maturity prior to engage in a big data initiative.

5.2 ORGANISATIONAL MODELS IN SMES FOR BUSINESS AND BIG DATA ANALYTICS

After assessing the maturity level, a company committed to invest in a "big data" program for improving its business performance faces the fundamental question of how to operationalise it, starting by the design of the required organisational structure. Here, a decision should be made regarding the degree of centralisation of analytics activities in support of business processes. There are three main organisational models that may be adopted: all data scientists in a single unit; a group of data scientists in each business unit; a hybrid solution called the "centre of excellence", that is, a unit from which distributed data scientists can obtain information and expertise as necessary. This framework is illustrated by Grossman and Siegel[51] and seems to be generally accepted, even though in the literature there is no specific analysis about which model could be better suited to a SME. A recent proposal was made by Rising et al.[52] to "create small self-organised and experimenting groups with a few members covering the key data scientist skills in a way to seek value adding benefits from big data with limited costs". However, a systematic take on this problem must also consider the SME's business complexity, how the strategic goals of the SME are shared among its units, the maturity of reporting and analytic capabilities, how information extracted from data has to be used (such as to build business vs leverage it as an advantage). The best combinations between organisational models and the mentioned features of the SMEs therefore need to be determined. A mechanism for this is yet to be developed.

A pattern that has been emerging is that large businesses are acquiring their skills by buying up smaller analytic and data handling firms or are forming strong alliances with these organisations. Most SMEs cannot form such alliances or do not wish to do so. Because SMEs cannot compete with big companies to recruit data scientists from the labour market, how to build the required skill internally is an important question also discussed by Rising et al.[52] in a class project report. One suggested approach is to use internal resources possessing knowledge of the business domain and to build their big data analytics capabilities in an inexpensive way through open source tools and

MOOCs. 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. Clearly, the scarcity of time that can be dedicated to learning is a major issue. A mechanism for assisting SMEs in this area is required once the areas of need have been identified,

The acquisition of big data analytic capabilities can also be regarded as a form of innovation. Innovation is mostly associated with Research & Development, aimed at obtaining new products or services for meeting (or creating) customer needs or keeping up with competitors. However, the ability of leveraging big data is also an innovation in the SME's business analytics and decision process. The open innovation concept offers an information exchange framework that can complement the SME's organisational model. It has been defined as 'the use of purposive inflows and outflows of knowledge to accelerate internal innovation, and to expand the markets for external use of innovation, respectively.' In particular, networking, a dimension of technology exploration, can fill in specific knowledge needs without having to spend large amounts of time and money to develop that knowledge internally[53].

For example, the matching of functional requirements and solutions offered by big data analytics products, referred to in Section 4, is an area that can greatly benefit from knowledge exchange about the most appropriate tools for addressing a given business problem. But how should such a network be built and maintained? Lee et al.[54] have examined the advantages of an intermediated network of SMEs for the commercialisation stage (a dimension of technology exploitation, in this case), where the intermediary could be a public body, for example, with the role of identifying appropriate collaborative partners, of constructing the network by appropriately matching SMEs and of managing the network. An indirect support that the intermediary can offer is to develop the culture of collaboration and to facilitate collaboration. Lee at al.[54] illustrate a positive intermediated network experience in Korea. It is to be seen whether this type of networking could be successfully used for developing big data analytics capabilities in SMEs either within the technology exploration or the technology exploitation dimension.

6 IMPLANTING A NEW MANAGEMENT METHODOLOGY – LESSONS FROM QUALITY MANAGEMENT

There is a pattern often repeated in the manner in which management revolutions --big changes-- appear and spread through the business world. Of course there are many variations, but the essence is frequently the same.

The revolution begins in large companies of a sector or sectors, when it starts to be clear that the new methodology works, it is adapted and adopted by other major companies in other sectors. From that moment it begins to appear on news, congresses and presentations in specialised forums: professional associations, trade magazines, chambers of commerce and the like. Almost in parallel it starts to be discussed and explained in business schools and offered by consulting companies, frequently as a must do panacea. It becomes a boom and this is when the more advanced SMEs begin to consider what they should do to get on the band wagon. Unfortunately what they frequently find is that the training provided in business schools and the services offered by consulting companies are not adapted to their characteristics, that they are designed for much larger companies. These pioneers find themselves alone in trying to implement the new methodology through a difficult process of trial and error. It is only when specific training and consultancy advice is developed that the new management methodology spreads among SMEs and becomes mainstream.

Let us consider as an example, even without going into any detail, the growth and spread of Total Quality Management. The story begins in the early 80s with some major companies (Ford, Xerox, HP, Philips, Bekaert, ABB ...) adopting quality improvement as a strategy. The first version of the ISO 9000 appears in 1987 and almost simultaneously the Malcolm Baldridge National Quality Award in the US and the European Quality Award in Europe. The latter promoted by the European Foundation for Quality Management (EFQM), created by 14 of the biggest European companies[55]. Remember that the true purpose of both awards was that their bases were to become models of TQM or Business Excellence -- as it started to be called-- used by organisations as a diagnostic and improvement tool.

Until the mid-nineties the adoption of TQM policies and practices was almost exclusively limited to large companies. It is from that point on that the activities to disseminate the methodology begin. In 1996 the DGIII of the European Commission produced a Quality Promotion Policy, in 1997 as part of it and jointly promoted by DGIII and the EFQM a book with 37 case studies of European companies are published[56]. At the same time, adaptations of the models to specific sectors (public sector, health, education ...) appear and immediately after to SMEs (the first was in 1999). Consultants as well as researchers and different government agencies start to produce and disseminate tools for self-assessment and maturity models based on the EFQM and Baldridge models. There are countless public events aimed to disseminate TQM tools and methods and the number of consultants dedicated to the topic continues to grow throughout the 90's and early 2000's, the same as the number of postgraduate and MBA's in TQM offered by universities and business schools. The European Regional Development Fund and European Social Fund funded a £1.25 million program of intervention aimed at implementing statistical process control, design of experiments and continuous improvement methods in SMEs. The funding was used to provide advice, training and dissemination. Where earlier unfunded efforts to access SMEs met with limited success, the 2 year campaign raised awareness and led to many SMEs adopting new practices[57].

It is difficult to say when TQM and its derivatives (Business Excellence, Lean Management, Six Sigma improvement ...) became methodologies and tools commonly used by SMEs. But it is clear, for example from the number of ISO 9001 certificates: 1.138.155 in 2014[58] that this is so.

It therefore seems clear that to accelerate the development and implementation of big data and analytics in SMEs, two important engines are training and the existence of accessible specialised advice. In fact a proper implementation of these two engines would attack simultaneously the problems and challenges posed in section 2: lack of understanding and dominance of domain specialists, shortage of in house data analytics expertise, bottlenecks in the labour market, lack of business cases, shortage of useful and affordable consultancy and lack of management and organisational models.

There are gaps in training both in the technological aspects, those related to database management, machine learning, data mining and knowledge of statistics and Operations Research, and in aspects of management. In other sections we have referred to the first aspect, that just a few years ago it was thought to be the bottleneck to a widespread of use of analytics and big data. Let us now have a brief comment with regard to the second, the importance of which, only very recently is starting to be realised. It is clear now, that a key to successful use of analytics is having people able to see where to use the analytical capabilities[59], the question is to have people who ask provocative questions, who can guide the search of technical people towards business interesting findings, who can interpret their analysis, who can think of interesting experiments to be run and at the end who are really data-based decision makers. Naturally these needs are more noticeable and are less covered in SMEs.

A prerequisite to both speed up the learning process and develop affordable consulting alternatives is to consolidate the knowledge about the best way to implement analytics into SMEs. This implies: collecting the knowledge already available, identifying success stories, finding the communalities, developing and trying methodologies and materials. It is necessary to synthesise the lessons into training materials and implementation steps, methods and tools.

What are then the alternatives to speed up the learning process in the SME environment?

In the long term, following the lessons of the TQM example discussed above it seems that the main alternatives lie in the university domain, the creation of programs specifically aimed at SMEs. Postgraduate and Master's programs for the more technical aspects and MBAs and management education programs for the managerial aspects. Eventually this will lead to incorporate those topics into existing degree courses. Some companies are looking really long term, for example SAS has developed a web-based tool for teaching data analytics to high school students, accessible at https://www.sascurriculumpathways.com/portal/

In the short term a quick development of the requisite technical and managerial skills needs the active support of Public Policies in the form of awareness campaigns involving public or semi-public organisations (chambers of commerce, professional associations, sector associations…), as well as active policies of subsidies or tax incentives. Putting together communities of practice and exchange forums, virtual as well as real will help to speed up the awareness and learning processes. A new and interesting possibility is the development of MOOC courses aimed specifically to develop the technical and managerial skills needed in SMEs.

The key to the creation of an affordable network of consulting capabilities is to develop a well-established and proved analytic system to help SMEs to successfully use analytics and big data. Following the TQM analogy, the availability of high level general models such as the ISO 9000 or the EFQM as well as particular models for particular sectors or particular parts of the TQM global effort, facilitated the appearance of an ever growing cohort of consultants. Most of them would not have dared enter the business without such guides. In turn, the competition between them caused on the one hand an improvement of their services and on the other hand a drop in their prices.

So, it is obvious that the development of a clear path to follow and tools to use will, in combination with the awareness and training measures, create a consultancy that is affordable and is a useful force. This could be complemented by services, on-site or on-line, provided by chambers of commerce and sectorial associations

7 OUTLOOK AND CONCLUSIONS

We have reviewed the challenges for SMEs faced with the need to tackle big data analytics to achieve the breakthroughs that large organisations are realising. The problems for the SME sector are complex, multifaceted, and transversal over various dimensions like IT, data analytic intelligence, organisational structure, managerial

models, capital structure and requirements, consulting, labour market, data security and legal aspects.

We have explored the IT requirements and state of the art solutions. Maturity models and organisational models for business and big data analytics have been discussed. We have reviewed the history of the spread of TQM as a blueprint for the management revolution that will be necessary to establish business analytics in the SME sector.

What is now needed is a mechanism to help SMEs to get started. We recommend a program of research to address the areas of concern that SMEs feel. An improved maturity model tailored to SMEs would be a first step for SMEs to find where they are on the path to big data analytics. To progress the construction of this model, there needs to be a review of SME attitudes as regards their process perspective, their functional perspective and their corporate adoption perspective. Table 1 shows the relationship between SME challenges and aspects of data maturity, however, more work needs to be done to produce guidelines for SMEs. There needs to be a mechanism for helping SMEs find the best combinations of organisational models, and a way to assist them to combine open innovation concepts with big data analytics capabilities as a way of acquiring data analytics skills.

Parallels with the adoption of new IT solutions and quality improvement initiatives such as TQM are helpful in determining how to assist SMEs. We can draw on previous successes in which massive intervention was funded on a national and European level as a way to bridge the gap between large organisations and SMEs. In terms of research and development, the complexity of the underlying problems requires a coordinated action involving several disciplines like IT, data analytics, economy and management science, jurisprudence, in close interaction with industry and business associations and policy makers. Enabling such an action will require public funding. Such funding is justified due to its previous success and the fact that SMEs are the powerhouse behind European employment and economic success.

REFERENCES

1. Sribar VT, Feinberg D, Gall N, Lapkin A, Beyer MA 2011. "Big Data" Is Only the Beginning of Extreme Information Management. Stamford, CT: Gartner Inc. Report ID number G00211490.

2. ABI Research 2013. Unlocking the Value of Big Data in Enterprises. *Application Analysis*; AN-1314.

3. Vesset D, McDonough B, Schubmehl D, Woodward A, Wardley M, Olofson CW. 2014. *Worldwide Business Analytics Software 2014–2018 Forecast and 2013 Vendor Shares*. Doc # 249926, International Data Corporation (IDC).

4. Woodward A. 2014. *Worldwide Advanced and Predictive Analytics Software 2014–2018 Forecast and 2013 Vendor Shares*. Doc # 249054, International Data Corporation (IDC).

5. KPMG 2015. *Mit Daten Werte schaffen - Report 2015*. KPMG AG: Zug, Switzerland; Bitkom Research: Berlin, Germany.

6. Provost F, Fawcett T. 2013. Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data* **1**(1): 51–59. DOI: 10.1089/big.2013.1508.

7. Haldane A. 2010. *Patience and Finance*. Oxford China Business Forum, Beijing, 2010.
http://www.bankofengland.co.uk/archive/Documents/historicpubs/speeches/2010/speech445.pdf [accessed 01 December 2015]

8. e-skills UK 2013. Big Data Analytics, Adoption and Employment Trends, 2012-2017. http://www.e-skills.com/Documents/Research/General/BigDataAnalytics_Report_Nov2013.pdf [accessed 21 December 2015]

9. TechNavio 2014. *Global SME Big Data Market 2014-2018*. TechNavio - Infiniti Research Ltd.

10. Vossen G, Lechtenbörger J, Fekete D. 2015. Big Data in kleinen und mittleren Unternehmen - eine empirische Bestandsaufnahme. Westfälische Wilhelms-Universität Münster, Germany. http://www.wi1.uni-muenster.de/pi/iai/publikationen/BigData.pdf [accessed 21 December 2015]

11. Russegger S, Freudenthaler B, Güntner G, Kieseberg P, Stern H, Strohmeier F. 2015. *Big Data und Data-driven Business für KMU*. Digital networked Data - Verein für Innovation und Erforschung vernetzter digitaler Daten, Salzburg, Austria.

12. Ghobakhloo M, Sabouri MS, Hong TS, Zulkifli N. 2011. Information Technology Adoption in Small and Medium sized Enterprises: An Appraisal of Two Decades Literature. *Interdisciplinary Journal of Research in Business* **1**(7): 53-80.

13. Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, Byers AH. 2011. *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.

14. Probst L, Frideres L, Demetri D, Vomhof B, Lonkeu OK. 2014. Customer Experience - Predictive analytics. http://docplayer.net/984399-Business-innovation-observatory-customer-experience-predictive-analytics-case-study-34-enterprise-and-industry.html#show_full_text [accessed 21 December 2015]

15. PayScale 2015a. Data Scientist, IT Salary. http://www.payscale.com/research/US/Job=Data_Scientist,_IT/Salary [accessed 13 September 2015]

16. PayScale 2015b. Software Engineer/Developer/Programmer Salary. http://www.payscale.com/research/US/Job=Software_Engineer_%2f_Developer_%2f_Programmer/Salary [accessed 13 September 2015]

17. Ahlemeyer-Stubbe A, Coleman SY. 2014. *A Practical Guide to Data Mining in Business and Industry,* Wiley: London, England.

18. Wadhwa K. 2014. BYTE: Big data roadmap and cross-disciplinary community for addressing societal Externalities, *European Data Forum,* Athens, Greece, 20 March 2014.

19. Felber S. 2011. *SMEs vs the consulting industry*, European CEO. http://www.europeanceo.com/business-and-management/smes-vs-the-consulting-industry/ [accessed 21 December 2015]

20. McAfee A, Brynjolfsson E. 2012. Big Data: The Management Revolution. *Harvard Business Review*, October 2012: 61-68.

21. Capgemini 2012. Measuring Organisational Maturity in Predictive Analytics: the First Step to Enabling the Vision. Capgemini, Paris. https://www.capgemini.com/resource-file-access/resource/pdf/Measuring_Organizational_Maturity_in_Predictive_Analytics__the_First_Step_to_Enabling_the_Vision.pdf [accessed 21 December 2015]

22. Schäfer A, Knapp M, May M, Voß A. 2012. *BIG DATA – Vorsprung durch Wissen*. Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS, Sankt Augustin, Germany.

23. Lacey D, James BE. 2010. *Review of Availability of Advice on Security for Small/Medium Sized Organisations*. Information Commissioner's Office: Wilmslow, UK.

24. Rains T. 2014. Cyber threats to Windows XP and guidance for Small Businesses and Individual Consumers. https://blogs.microsoft.com/cybertrust/2014/03/24/cyber-threats-to-windows-xp-and-guidance-for-small-businesses-and-individual-consumers/ [accessed 01 December 2015]

25. Edwards G. 2015. Cyber Security Threats and Data Loss: Hacker Attacks on SMB Organizations. http://blog.watchpointdata.com/cyber-security-threats-and-data-loss-hacker-attacks-on-smb-organizations/ [accessed 01 December 2015].

26. Viuker S. 2015. Cybercrime and hacking are even bigger worries for small business owners. The Guardian. http://www.theguardian.com/business/2015/jan/21/ cybersecurity-small-business-thwarting-hackers-obama-cameron [accessed 01 December 2015]

27. European Union Agency for Fundamental Rights, Council of Europe 2014. *Handbook on European data protection law*, Publications Office of the European Union: Luxembourg. DOI: 10.2811/69915

28. Davidsson P. 1989. Entrepreneurship – and after? A study of growth willingness in small firms. *Journal of Business Venturing* 1989; **4**: 211–226.

29. Pissarides, F. 1999. Is lack of funds the main obstacle to growth? ebrd's experience with small- and medium-sized businesses in central and eastern Europe. *Journal of Business Venturing* **14**(5): 519–539. DOI: 10.1016/S0883-9026(98)00027-5

30. Bartlett W, Bukvić V. 2001. Barriers to SME Growth in Slovenia. *Economic Policy in Transitional Economies* 05/2001; **11**(2): 177-195. DOI: 10.1023/A:1012206414785

31. Fuller-Love N. 2006. Management development in small firms. *International Journal of Management Reviews* **8**(3): 175-190. DOI: 10.1111/j.1468-2370.2006.00125.x

32. Muthaih K, Venkatesh S. 2012. A study on the barriers affecting the growth of small and medium enterprises in India. *International Journal of Research in Computer Application Management* **2**(1): 77–81.

33. Zhong B. 2008. Economic analysis on the financing difficulty of small and medium enterprises based on the perspective of information asymmetry. *Economist* **1**: 225-226.

34. Muller P, Gagliardi D, Caliandro C, Bohn NU, Klitou D. 2014. *Annual Report on European SMEs 2013/2014 - A Partial and Fragile Recovery*. European Commission.

35. Statistisches Bundesamt 2012. Kleine & mittlere Unternehmen (KMU), Mittelstand. https://www.destatis.de/DE/ZahlenFakten/GesamtwirtschaftUmwelt/UnternehmenHand

werk/KleineMittlereUnternehmenMittelstand/KleineMittlereUnternehmenMittelstand.ht
ml [accessed 13 September 2015]

36. European Commission 2014. Helping SMEs fish the Big Data Ocean.
http://cordis.europa.eu/result/rcn/93077_en.html [accessed 26 July 2015]

37. Kiron D, Prentice PK, Ferguson RB. 2012. Innovating With Analytics. *MIT Sloan
Management Revue*, **54**(1): 47-52

38. Labrinidis A, Jagadish HV. 2012. Challenges and Opportunities with Big
Data. *Proceedings of the VLDB Endowment* **5**(12): 2032-2033. DOI:
10.14778/2367502.2367572

39. Cattell R. 2011. Scalable SQL and NoSQL data stores. *ACM SIGMOD Record*
39(4): 12–27.

40. Borthakur D. 2007. *The Hadoop Distributed File System: Architecture and Design.*
The Apache Software Foundation.
https://svn.apache.org/repos/asf/hadoop/common/tags/
release-0.17.0/docs/hdfs_design.pdf [accessed 21 December 2015]

41. Lakshman A, Malik P. 2009. Cassandra: structured storage system on a P2P
network. *Proceedings of the 28th ACM symposium on Principles of distributed
computing* (PODC '09), **5**. ACM: New York, NY. DOI:10.1145/1582716.1582722

42. De Candia G, Hastorun D, Jampani M, Kakulapati G, Lakshman A, Pilchin A, Siva
subramanian S, Vosshall P, Vogels W. 2007. Dynamo: Amazon's Highly Available
Key-value Store. *Proceedings of the twenty-first ACM SIGOPS symposium on
Operating systems principles* (SOSP '07), 205-220. ACM: New York, NY. DOI:
10.1145/1294261.1294281

43. Chodorow K. 2013. *MongoDB: the definitive guide*, O'Reilly Media Inc:
Sebastopol, CA.

44. Anderson JC, Lehnardt J, Slater N. 2010. *CouchDB: The Definitive Guide*. O'Reilly Media Inc: Sebastopol, CA.

45. George L. 2011. *HBase: the definitive guide*. O'Reilly Media Inc: Sebastopol, CA.

46. Dean J, Ghemawat S. 2008. Mapreduce: simplified data processing on large clusters. *Communications of the ACM* 2008; **51**(1): 107–113.

47. Mell P, Grance T. 2011. The NIST definition of cloud computing. *NIST Special Publication* 800-145, National Institute of Standards and Technology, US Department of Commerce.

48. Zhang Q, Cheng L, Boutaba R. 2010. Cloud computing: state-of-the-art and research challenges. *Journal of Internet Services and Applications* **1**(1): 7-18.

49. Harvey C. 2012. 50 Top Open Source Tools for Big Data. http://www.datamation.com/print/http://www.datamation.com/data-center/50-top-open-source-tools-for-big-data-1.html [accessed 21 November 2015]

50. Loshin D. 2015. Comparing the leading big data analytics software options. http://searchbusinessanalytics.techtarget.com/feature/Comparing-the-leading-big-data-analytics-software-options. [accessed 21 December 2015]

51. Grossman RL, Siegel KP. 2014. Organisational models for big data and analytics. *Journal of Organization Design*, **3**(1): 20-25. DOI: 10.7146/jod.9799

52. Rising CJ, Kristensen M, Tjerrild-Hansen S. 2014. Big Data too Big for SMEs? http://web.stanford.edu/class/msande238/projects/2014/GainIT.pdf [accessed 20 September 2015]

53. Van de Vrande V, de Jong JPJ, Vanhaverbeke W, de Rochemont M. 2009. Open Innovation in SMEs: Trends, Motives and Management Challenges. *Technovation* **29** (6-7): 423-437. DOI:10.1016/j.technovation.2008.10.001

54. Lee S, Park G, Yoon B, Park J 2010. Open innovation in SMEs - An intermediated network model. *Research Policy* **39**: 290-300.

55. Conti T. 2007. A history and review of the European Quality Award Model. *The TQM Magazine* 03/2007; **19**(2): 112-128. DOI: 10.1108/09544780710729962

56. Hadjorno TW, ten Have S, ten Have WD. 1997. *The European way to excellence: how 35 European manufacturing, public and service organisations make use of quality management*. Directorate General III Industry, European Commission.

57. Stewardson DJ, Coleman SY 2003. *Success and failure in helping SMEs, a three-year observational study*. Industry and Higher Education 03/2003; **17**(2):125-130. DOI: 10.5367/000000003101296783

58. International Organisation for Standardization 2014. *The ISO Survey of Management System Standard Certifications - 2014*. International Organisation for Standardization, Geneva. http://www.iso.org/iso/iso_survey_executive-summary.pdf?v2014 [accessed 21 December 2015]

59. Ransbotham S, Kiron D, Prentice PK. 2015. The Talent Dividend: Analytics talent is driving competitive advantage at data-oriented companies. *MIT Sloan Management Review*, April 2015.

Table 1. Cloud computing infrastructures

		SaaS: Software as a Service	PaaS: Platform as a Service	IaaS: Infrastructure as a Service
Features		Capability to access software running on servers maintained and deployed by third parties.	An environment in which the operating system and server software, as well as the underlying server hardware and network infrastructure are managed by the vendor.	A highly automated and scalable pool of computing resources, complemented by cloud storage and network capabilities.
Advantages		Reduced cost of software ownership, such as licencing and technical maintenance.	User only manages specific functionalities of interest.	Instead of having to purchase hardware, users can purchase IaaS based on consumption, similar to electricity or other utility billing.
Management of Computing Component	**Application**	*Vendor*	*User*	*User*
	Data	*Vendor*	*User*	*User*
	Runtime	*Vendor*	*Vendor*	*User*
	Middleware	*Vendor*	*Vendor*	*User*
	Operating System	*Vendor*	*Vendor*	*User*
	Virtualization	*Vendor*	*Vendor*	*Vendor*
	Servers	*Vendor*	*Vendor*	*Vendor*
	Storage	*Vendor*	*Vendor*	*Vendor*
	Networking	*Vendor*	*Vendor*	*Vendor*

Table 2. The relationship between the challenges SMEs are facing and the dimensions characterising their maturity as organisations that make effective use of data resources.

CHALLENGES	DIMENSIONS						
	Business Strategy	Data Manag.	Spec. people & skills	Tech. infra.	Enterprise adopt.	Leadership and corp. culture	Data gover.
1.Lack of understanding			***			**	
2.Dominance of domain specialists			***			**	
3.Cultural barriers and intrinsic conservatism			**		***	***	
4.Shortage of in-house data analytic expertise			***		***		
5.Bottlenecks in the labour market			***				
6.Lack of business cases	**					**	
7.Shortage of useful and affordable consulting and business analytics services		**	**	***			
8.Nontransparent software market			**				
9.Lack of intuitive software		**	**	**			
10.Lack of management and organisational models	***				***	**	
11.Concerns on data security						*	***
12.Concerns on data protection and data privacy						*	***
13,Different venture concept					**	**	
14.Financial barriers		**	**	***			

Legend: strength of the relationship/impact: * - low; ** - moderate; *** - high.

Biographies

Shirley Coleman:

Shirley Coleman, PhD is Principal Statistician at the Industrial Statistics Research Unit, Newcastle University and a visiting scholar at the Faculty of Economics, Ljubljana University, Slovenia. She works on data analytics in the UK gas energy sector and contributed a highly ranked impact case study to Newcastle University's 2014 Research Excellence Framework. She is PI on 3 two-year Innovate UK funded knowledge transfer partnerships applying data analytics in SMEs. She is a past President of the European Network for Business and Industrial Statistics (ENBIS), an elected member of the International Statistics Institute and a Chartered Statistician of the Royal Statistical Society, leading on developing relations with Operations Research and open data. She is an active reviewer and conference organiser to help address the challenge of communication and dissemination to the wider community.

Rainer Goeb

Rainer Goeb is a Professor of Statistics at the Institute for Applied Mathematics and Statistics, University of Wurzburg, Germany. His research interests and industrial consulting activities cover statistical sampling, statistical quality control, statistics in accounting and auditing, and predictive analytics. He is currently an associate editor of QTQM (Quality Technology and Quantitative Management), and of the Journal of Methodological and Applied Statistics. He is the convenor of the subcommittee SC 5 `Acceptance Sampling' of the ISO Technical Committee 69 `Applications of statistical methods', and has been president of the European Network for Business and Industrial Statistics (ENBIS) in 2014 and 2015. He is the founder and scientific head of prognostica consulting GmbH, a company specialised in predictive analytics.

Giuseppe Manco:

Giuseppe Manco is a researcher at the Institute of High Performance Computing and Networks (ICAR-CNR, Italy) and a contract professor at University of Calabria, Italy. His current research interests include knowledge discovery and data mining, Recommender systems and Social Network analysis. He has been serving in the committees of several conferences on knowledge Discovery and Data Mining, including IEEE ICDM, ECMLPKDD, SIAM SDM. He has been serving as a guest editor for the Journal of Intelligent Information Systems and as a member of the editorial board of the ECML/PKDD 2014 Journal track. He was recipient of the 2014 Yahoo Faculty Research and Engagement Program. He is the founder of Open Knowledge Technologies S.r.l. and the scientific advisor of Exeura S.r.l., two companies specialised in Big Data analytics.

Antonio Pievatolo

Antonio Pievatolo, PhD in Statistics, is a researcher at the Institute for Applied Mathematics and Information Technology "Enrico Magenes" (IMATI-CNR, Italy) and a contract professor for the PhD program in Statistics at Bocconi University. He is President-elect of the European Network for Business and Industrial Statistics (ENBIS). He is an associate editor of Applied Stochastic Models in Business and Industry. His research focusses on problems in engineering and technology using stochastic modelling tools, with a preference for Markov and point processes. His main contributions can be found in the application of stochastic modelling to reliability and to electrical power systems.

Marco Reis

Marco Reis, PhD in Chemical Engineering, Professor in the Department of Chemical Engineering of the University of Coimbra, Portugal. He is currently President of the European Network for Business and Industrial Statistics (ENBIS), President of PRODEQ – Association for the Development of Chemical Engineering and responsible for the Process Systems Engineering (PSE) research group at the department of Chemical Engineering. He lectures courses on process systems engineering, quality technology and management, management and entrepreneurship, and process improvement. His research interests are centred on the field of process systems engineering (system identification, fault detection and diagnosis, control and optimization), statistical process control of complex large-scale processes, data-driven multiscale modelling, chemometrics, design of experiments and industrial statistics. Other areas of interest include multivariate image analysis, systems biology and process improvement through initiatives such as six-sigma and lean manufacturing.

Xavier Tort-Martorell

Xavier Tort-Martorell is Professor of Statistics at Universitat Politècnica de Catalunya. Barcelona TECH. His research interests and industrial consulting activities cover quality management and improvement, six sigma, design of experiments and data based decision making. He is currently an associate editor of TQM&BE (Total Quality Management & Business Excellence), QTQM (Quality Technology and Quantitative Management), and of the Int. Journal of Quality and Service Sciences. Has been assessor of the European Quality Award, member of the jury of the Catalan Quality Prize and president of the European Network for Business and Industrial Statistics (ENBIS) in 2012 and 2013. He is the founder Datancia, a company specialized in data based decision making.