# How to Engrain a Big Data Mindset into Our Managers' DNA: Insights from a Big Data Initiative in a French Business School

**Chapter** · January 2016

DOI: 10.4018/978-1-4666-9649-5.ch005

CITATION
1

READS
136

**1 author:**

Kevin Daniel André Carillo
Toulouse Business School
**49** PUBLICATIONS **112** CITATIONS

SEE PROFILE

# Managing Big Data Integration in the Public Sector

Anil Aggarwal
*University of Baltimore, USA*

A volume in the Advances in Public Policy and Administration (APPA) Book Series

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA, USA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com



Library of Congress Cataloging-in-Publication Data

CIP Data Pending
ISBN: 978-1-4666-9649-5
eISBN: 978-1-4666-9650-1

This book is published in the IGI Global book series Advances in Public Policy and Administration (APPA) (ISSN: Pending; eISSN: Pending)

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

# Chapter 5
# How to Engrain a Big Data Mindset into Our Managers' DNA:
## Insights from a Big Data Initiative in a French Business School

**Kevin Daniel André Carillo**
*Toulouse University, France & Toulouse Business School, France*

## ABSTRACT

*The industry experts' claim that big data will be the next frontier for innovation, competition, and productivity, has turned from an unfounded omen to an obvious reality. In the looming data-driven business era, data will pervade organizations and will thus affect their nature and functioning. The central role played by data is about to impact the main mechanism that fuels the life of organizations, decision making, which will then gradually redefine the job and function of managers. In May 2014, a French top-ten business school launched a 2-week big data initiative aiming at immersing future managers into a big data simulation and also sensitizing them to the overall cultural shift that big data is about to engender within organizations. Using semi-real data about a French newspaper, 100 graduating students were engaged into a big data serious game which main objective was to use data visualization techniques on a large dataset hosted on a Hadoop cluster in order to derive business recommendations to higher management. By conducting observations, taking field notes, and interviewing participants, important insights could be derived for educational institutions that are contemplating the importance of including big data into their curricula but also for organizations that strive to engrain a big data mindset within their managers.*

## INTRODUCTION

There is now no doubt that there exists a reality beyond the big data buzz. By realizing that data have the potential to be a key business driver, organizations are gradually shifting towards a new business era in which data will be treated as precious and strategic resources. McKinsey's 2011 claim (Manyika

DOI: 10.4018/978-1-4666-9649-5.ch005

et al., 2011) about big data being the next frontier for innovation, competition, and productivity, has now turned from omen to an understated reality. In this data-driven business era, data will pervade organizations and will thus affect their nature and functioning. As a consequence, the main mechanism that fuels the life of organizations, decision making, will be significantly impacted by the central role played by data. In other words, big data is gradually redefining the job of managers by incorporating data into decision processes.

In May 2014, a French business school launched a 2-week big data initiative aimed at immersing future managers into a big data simulation and engraining into them a big data mindset. Using semi-real data about a French newspaper, 100 graduating students were engaged into a big data serious game which main objective was to use data visualization techniques in order to derive business recommendations from user data hosted on a Hadoop cluster. By conducting observations and interviewing students during the serious game, important lessons could be derived for organizations that strive to engrain a big data mindset within their managers. For instance, the workshop clearly highlighted that the lack of solid database-related skills as well as certain statistical concepts could prevent managers from effectively extracting the business value contained into data. Besides, lessons generated from the workshop can also help business schools and universities to identify pedagogical strategies to sensitize future managers towards the predominant role that big data is going to play as well as to provide them with the necessary skills and mindset.

The chapter particularly targets business schools that are contemplating the decision to integrate big data into their mainstream curricula as well as into specific trainings. It also provides insights about how big data is gradually redefining the function and job of managers. By doing so, this chapter also aims at helping organizations to tailor effective trainings to raise awareness about the transformation of management in the big data era. The public sector, perceived as "the most fertile terrain" for benefitting from entering the data-driven era, is particularly targeted as governments have collected huge amounts of 'siloed' data and have for long suffered from an overall lack of efficiency and performance (Brown et al., 2011).

This book chapter is organized as follows. In a first section, it explains that beyond the big data buzz there exists a big data reality that will globally impact all businesses and our society as a whole. It concludes by contemplating how big data will affect managers, how business schools shall take it into consideration, and how the Information Systems field can play a leading role in developing and delivering adapted curricula. After succinctly presenting the French business school context, the workshop and details about its pedagogical engineering are presented. Based on results drawn from data collected throughout the event, the last section derives insights and recommendations for organizations as well as business schools and universities. A conclusion summarizes the key points raised throughout the chapter.

## BIG DATA, BIG CHANGE, BIG SKILL SHORTAGE

### The Quest for a 'Good' Big Data Definition

Attempts to determine the origin of the term 'big data' have led to inconclusive results. However, what is sure is that the "big" part of "big data" is ill-chosen as it implies ideas of numbers or quantities omitting more predominant aspects such as its strategic importance or its business relevance. As a result, the connotations around the use of 'big data' tends make companies think in terms of *how big do my data*

*need to be to do some big data*? … whereas the key question shall rather be *what strategic value can I derive from my data no matter its size?*

Big data definitions abound as big data experts have attempted to delineate the blurry boundaries that surround the big data phenomenon. Realizing that the shared origin of big data among academia, industry and the media had led to ambiguous and contradictory definitions, Jonathan Stuart Ward and Adam Barker (2014) of University of St Andrews (UK) attempted to close the debate by collating the various definitions having gained some consensus and providing a clear and concise definition. Table 1 presents the big data definitions compiled by Ward and Barker as well as other definitions that have been commonly acknowledged in the industry and education.

*Table 1. Most commonly acknowledged big data definitions*

| Source / Author(s) | Big Data Definition |
|---|---|
| *Merriam Webster* | An accumulation of data that is too large and complex for processing by traditional database management tools. |
| *Oxford Dictionaries* | Extremely large data sets that may be analysed computationally to reveal patterns, trends, and associations, especially relating to human behaviour and interactions:<br>much IT investment is going towards managing and maintaining big data |
| *Gartner's glossary (2014)* | 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. |
| *IDC* | Big Data technologies as a new generation of technologies and architectures designed to extract value economically from very large volumes of a wide variety of data by enabling high-velocity capture, discovery, and/or analysis. This definition encompasses hardware, software, and services that integrate, organize, manage, analyze, and present data that is characterized by "four Vs" — volume, variety, velocity, and value (discussed in the sections that follow). |
| *Gartner (2012)* | In 2001, a Meta (now Gartner) report noted the increasing size of data, the increasing rate at which it is produced and the increasing range of formats and representations employed. This report predated the term "dig data" but proposed a three-fold definition encompassing the "three Vs": Volume, Velocity and Variety. This idea has since become popular and sometimes includes a fourth V: veracity, to cover questions of trust and uncertainty. |
| *Oracle* | Big data is the derivation of value from traditional relational database-driven business decision making, augmented with new sources of unstructured data. |
| *Forrester* | The practices and technology that close the gap between the data available and the ability to turn that data into business insight. |
| *Intel* | Big data opportunities emerge in organizations generating a median of 300 terabytes of data a week. The most common forms of data analyzed in this way are business transactions stored in relational databases, followed by documents, e-mail, sensor data, blogs, and social media. |
| *Microsoft* | Big data is the term increasingly used to describe the process of applying serious computing power—the latest in machine learning and artificial intelligence—to seriously massive and often highly complex sets of information. |
| *The Method for an Integrated Knowledge Environment open-source project* | The MIKE project argues that big data is not a function of the size of a data set but its complexity. Consequently, it is the high degree of permutations and interactions within a data set that defines big data. |
| *The National Institute of Standards and Technology* | NIST argues that big data is data which "exceed(s) the capacity or capability of current or conventional methods and systems." In other words, the notion of "big" is relative to the current standard of computation. |
| *Jonathan Stuart Ward and Adam Barker* | Big data is a term describing the storage and analysis of large and or complex data sets using a series of techniques including, but not limited to: NoSQL, MapReduce and machine learning. |

It is striking that nearly all definitions suffer from the same tendency to reduce big data to a technological change in the global economic landscape, rather focusing on the "big" part of "big data". Besides, mentions of the human side of big data: business, organizations, and people are rarely present in the definitions. Even Ward and Barker's definition, which is supposed to be a meta-definition, thus hypothetically more encompassing, is entirely focused on the technical side of big data: *the storage and analysis of large and or complex data sets using a series of techniques including, but not limited to: NoSQL, MapReduce and machine learning*. If big data is such a major phenomenon, what is the societal impact of the use of tools and technologies such as Hadoop, HBase, BigTable, Hive, or NoSQL?

## Big Data Is Real

'Selfie' was chosen by the Oxford Dictionaries as the 2013 word of the year as the frequency of the word had increased by 17,000% over the year. If there were an equivalent contest with business terms, there is little doubt that 'big data' would be the winner by far. Indeed, it is nowadays nearly impossible to go a day without encountering some mention of the term 'big data' in the corporate sphere or in the media. Nonetheless, there is a business reality beyond the big data buzz. Since McKinsey's 2011 report (Manyika et al., 2011) that took part in drawing attention on the big data phenomenon, companies are multiplying efforts and investments on big data initiatives. A 2012 Gartner report predicted that IT spending on big data in the U.S. would hit $34 billion and concluded that 64% of the surveyed companies had made such investments (Gartner, 2012a). Previsions for the coming years are even higher. IDC has predicted that the big data market (including infrastructure, services, and software) would reach $16.1 billion in 2014 (Gil Press, 2013) (a growth 6 times faster than the growth of the overall IT market). Behind such a frenzy on investing into big data technologies, big data has a reality that goes beyond the mere idea of a technological revolution. It is a real shift towards a new business era in which data play the central role.

## Big Data Talent Shortage for Managers

It is now clearly acknowledged that the global gap between the demand for big data talent and the supply of talent is one of the key challenges that seriously hamper big data implementations across organizations. A 2012 Gartner report predicted that by 2015, the demand for big data related jobs would reach 4.4 million jobs throughout the world while barely a third of those jobs would be filled (Gartner, 2012b). Predictions for the following years are not any better as the gap is foreseen to significantly widen…

Such prediction shot an alarm signal for all universities and business schools, urging them to develop new programs and adapt older ones to urgently answer the demand from organizations. Business schools may be tempted to think that this drastic skill shortage only concerns IT-related jobs. Such a view is absolutely erroneous as it would simply omit the human layer that surrounds the big data technological revolution, and even the entire 'data-transformation' that organizations are about to live.

Peter Sondergaard's words (senior vice president at Gartner and global head of Research) during his 2012 keynote speech of the Gartner Symposium/ITxpo, illustrate how the boundaries of the skill shortage will bridge the IT sphere:

*In addition, every big data-related role in the U.S. will create employment for three people outside of IT, so over the next four years a total of 6 million jobs in the U.S. will be generated by the information economy (Sondergaard, 2012).*

Meanwhile, Gartner (2012a) predicted a lack of 1.5 million managers and analysts with the skills to understand and make decisions based on the analysis of big data. The reports also foresees a lack of an additional 1.5 million managers and analysts in the United States for skills related to the effective identification and formulation of big data opportunities, and the consumption of their associated analysis. The accuracy of the above numbers is perhaps questionable but it is undoubtable that business schools are in the same boat and must urgently act. Training future managers being the main mission of business schools…

## A New Business Paradigm where Data is Digital Oil

The main argument of this book chapter is that big data heralds a major shift towards a new business paradigm that will impact every aspect of organizations. The direction of the link between business and data is gradually shifting. Indeed, data used to be there to support business. Now, organizations have understood that data have the potential to 'drive' business by providing new opportunities, strategies, allowing new business models, and transforming most business processes. Data are becoming the new digital oil that runs through the veins of organizations and that must be cherished in order to keep it pure (accurate and reliable) and valuable. Beyond the technological buzz surrounding the term 'big data' lays a much broader truth: the advent of a new data-driven business paradigm. Back in 2004, Yahoo! was among the first to appoint a Chief Data Officer (Usama Fayyad), this was already announcing the beginning of this new era. Very few people had realized it at that time…

A few years ago, when some organizations started realizing the potential of analyzing big data, they thought that a clever and talented individual could alone unleash the power of big data and do wonders. The quest for the mythical five-legged sheep that master statistics, computer science, and business, started; making the data scientist job one of the sexiest in the world. There is no doubt that there exists extremely talented individuals whose profile is close to the data scientist ideal. Nonetheless, examples showing that the 'data scientist can do it all' model does not work, abound. In this new data-driven business world, it is not one person that must be different but it is rather organizations as a whole and their overall functioning that must change. Data must become an organizational culture rather than technical means that support businesses. Data must become the center of attention for all employees from any hierarchical level as everybody in an organization takes part in the life cycle of data from its generation, storage, to its analysis and consumption. In retail and distribution companies, creating such a 'data culture' starts by making cashiers and employees working in warehouses aware that data are important and precious as these individuals play a crucial role by being at the very beginning of the data lifecycle.

## New Skills or New DNA?

Taking into consideration the above arguments, the big data skill shortage shall be seen from two complementary angles. First, such shortage can be seen *vertically*, as big data and the societal phenomenon that surrounds it have engendered new skills and created new functions in organizations.

There is a wealth of new technologies and tools around big data. Highest paying IT jobs ask for specific skills in NoSQL, Apache Hadoop, Python, and even MongoDB, HBase, Hive … Besides, starting from the function of Chief Data Officer, an array of new big data-related jobs have appeared in organizations such as big data scientist, big data analyst, big data visualizer, big data manager, big data solutions architect, big data engineer, big data consultant… to name but a few. However, by consider-

ing that behind the big data frenzy hides the looming shift towards a data-driven business era, the big data skill shortage goes way beyond the figures compiled by experts' reports. In this new data-driven business era, big data must become a mindset that is engrained into all employees' DNA. This is all the more important for managers from any hierarchical level as no matter their decisional scope (top, middle or operational), data and big data is becoming an inherent part of the inner mechanisms that govern the functioning of organizations: *decision making*. This is the perspective that is adopted throughout the remainder of the book chapter.

## Information Systems Coming to the Rescue of Management Education

Since the alarming 2011 McKinsey report about the looming worldwide shortage in big data / data analytics skills, universities have launched tremendous efforts to improve their existing degree programs and create new offerings. Overall, there has been a worldwide tendency to: (1) Add a Business Analytics/Big Data/Data Science focus to existing MBA programs; and (2) Develop new Master's programs (mainly Master of Science programs) in Business Analytics/Big Data/Data Science. In the United States, the number of MS Analytics programs (combining MSA: Master of Science in Analytics, MSBA: Master of Science in Business Analytics, and MSDS: Master of Science in Data Science) has grown from less than 20 in 2012, to about 30 in 2013, up to more than 60 in 2014 and nearly 80 as of 2015 (Schoenherr and Speier-Pero, 2015). A review of the top MS in Analytics programs performed in 2013 revealed that the content of such programs was split across the three core areas: analytical and modeling tools (44% of the overall content on average), business processes and decision making (24%), and data management (23%) (Schoenherr and Speier-Pero, 2015). There is no doubt that the programs that have been created to answer the big data skill shortage have high academic and practical value. However, it is puzzling to realize that on average only 13% of the delivered content (ranging from 6% to 20%) integrates the three knowledge domains through the realization of "hands-on" projects. The new data-driven business paradigm engenders the fusion of the three domains. As a result, preparing future managers to evolve in such new business environment shall not be by concatenating courses of each of the three fields but rather by merging them in nearly every program unit. Only then, universities will be in a position to claim that they are efficiently preparing managers that will be capable of evolving in the data-driven business paradigm. The academic boundaries between disciplines have no more reason to hold. Big data is the result of the fusion of all three.

The main quest of the Information Systems discipline is to tackle challenges and identify opportunities that can have a long lasting scientific and societal impact (Chen, 2011). Shortly after the publication of the 2011 McKinsey report, the IS field realized the urgent need to revisit existing curricula and launch action plans to provide Business Intelligence (BI) and Business Analytics (BA) education programs that would address the new generation of data/analytics savvy and business students and professionals (Chen et al., 2012). In 2012, the AIS Special Interest Group on Decision Support, Knowledge and Data Management Systems (SIGDSS) and the Teradata University Network (TUN) cosponsored the Business Intelligence Congress 3 conducted surveys to assess academia's response to the growing market need for students with Business Intelligence (BI) and Business Analytics (BA) skill sets with an emphasis on 'big data' (Wixom et al., 2014). The report concluded that the IS field was particularly well positioned to train the next-generation BI/BA workforce.

The most challenging curricular aspects when reflecting on the most efficient means to prepare future managers, is the breadth and depth of skill sets (at the junction of the three expertise domains) that are

needed to become a highly capable professional (Schoenherr & Speier-Pero, 2015). Since its early beginnings some 30 to 40 years back, Information Systems (IS) has been an interdisciplinary field in nature and has mastered on trans-disciplinary dialog (Galliers, 2003). If one discipline could claim legitimacy in being the perfect candidate to deliver education programs based on a new knowledge domain that is at the junction of computer science (data/database management), statistics (analytical and modeling tools), and business (business processes and decision making), Information Systems is the first discipline that shall immediately come to mind. The advent of the data-driven business era is an unprecedented opportunity for IS departments to play a central role by leading the education of next-generation professionals.

This chapter also aims at warning organizations that following the big data hive will go beyond gaining big data skills and creating big data functions. The remainder of this chapter is a testimony of a big data initiative that took place at Toulouse Business School in 2014 which main objective was to attempt to engrain such big data mindset into the DNA of future managers. This chapter does not have the pretension to provide lessons to other educational institutions. Indeed, over the last few years, some business schools and universities throughout the world have started launching an array of quality big data programs and trainings. This chapter particularly targets business schools that are in the initial phase of contemplating the relevance to teach big data to future managers.

## PRESENTATION OF THE 'BIG DATA BETTER DECISIONS' WORKSHOP

The Big Data Better Decisions workshop was a two-week event that took place at the end of the 2013-2014 university year of the Master in Management program. It involved 100 participants and twelve corporate partners that took part to the different sessions. The objectives of the workshop were to sensitize future managers to the reality of big data, to train them to a big data visualization tool, and to participate in a serious game aiming at simulating a business task in a big data context.

### Specificities of French Business Schools

Business schools in France are based on the model of French *grandes écoles*. They are 'elite' higher education entities that lay outside the French university system. *Grandes écoles* ((usually specialized in either engineering, sciences, or business) select students based on national rankings that are established through competitive written and oral exams. Most candidates complete two years of dedicated preparatory classes before taking the different *grandes écoles'* entrance exams even though more and more students are accepted through a parallel admission process (involving an ever increasing number of foreign students). The workshop presented in this book chapter was organized at one of the top ten French business schools that is representative of the French business school system.

### Participant Profile

Because the program aimed at sensitizing future managers to the global shift into a data-centric business paradigm, it was important to select participants that represented a broad spectrum of skills and expertise and that had some work experience. Students in the final year of the Master of Management program of the school seemed to be ideal targets. Besides, we selected only students that were following the apprenticeship track of the program which guaranteed a minimum of one to two years of work experience.

All the majors of the business school were represented (marketing, finance, accounting, international business, entrepreneurship, and supply chain management) as well as nearly all possible specializations. The participants, aged between 22 and 25 years old, worked for all types of organizations starting from small and medium-sized companies to large multinational corporations in a variety of sectors such as banking, aerospace, retail, distribution, energy, oil… as well as consulting firms. The workshop was designed for big data neophytes. The degree of knowledge about big data was close to null and limited to having read a few newspaper articles. No single student had worked directly or indirectly on big data issues in their respective company.

## Workshop Structure and Content

The main structure of the workshop is presented in Table 2. The first week of the workshop consisted of four half-day sessions on a range of complementary big data topics (see Table 3 for the complete list of the big data sessions and the associated learning objectives), a day session on presenting the serious game and the business concepts surrounding it (social network analysis and customer/user experience data), and two full days for training sessions aiming at mastering the data visualization tool that had to be used for the serious game. A certificate was delivered to participants having demonstrated a good mastery of it by successfully passing two online quizzes. The second week involved four half-day sessions with big data experts, the rest of the time being dedicated to partaking the serious game and working on the case. Three professors and an admin person were there to coach the groups and help with technical problems. Participants also used an online forum to ask questions to the professionals that had taken part in creating the serious game and who had a good understanding of the case and the dataset.

## Serious Game: Data Visualization Business Case

Participants were placed into groups of five, ensuring that each group had members from different domains of expertise and work experience in distinct sectors/industries. A full-day session was dedicated to present the business case associated with the serious game as well as the business concepts required to get an in-depth grasp of the specificities of the case. The groups were provided four days to work on the case. The final day consisted of attending the group presentations and evaluating their performance. Four big data experts (who had knowledge about the case and its business context) were invited to partake the final stage of the event.

*Table 2. Big data better decisions – Workshop structure*

| WEEK 1 | Monday | Tuesday | Wednesday | Thursday | Friday |
|---|---|---|---|---|---|
| Morning | Sessions with big data professionals/ specialists | | Serious game presentation | Training/certification | |
| Afternoon | | | | Data visualization tool | |
| WEEK 2 | Monday | Tuesday | Wednesday | Thursday | Friday |
| Morning | Sessions with big data professionals/specialists | | | | Student presentations and award ceremony |
| Afternoon | Serious game (group work and coaching) | | | | |

*Table 3. Big data sessions: Title, content, and speaker(s)*

| Session | Theme | Learning Objective(s) | Content | Speaker |
|---|---|---|---|---|
| **1** | **Introduction to big data** | ***LO1:*** Gain awareness about the advent of the data-driven business paradigm.<br>***LO2:*** Understand the nature, size and scope of the big data phenomenon<br>***LO3:*** Distinguish big data buzz vs. reality | • Defining big data<br>• The three/five/seven Vs<br>• Beyond the Vs, a business paradigm<br>• Famous big data stories and applications | Information System assistant professor (specialized in big data) |
| **2** | **Big data challenges, issues, and applications** | ***LO4:*** Understand how the different sectors and industries are (about to be) impacted/ transformed. | • What are the main issues, challenge, and applications of big data?<br>• Sectors and industries impacted by big data<br>• Feedback – Project in online retail industry | Big data consultant / CEO of big data startup |
| **3** | **Story of a big data startup** | ***LO5:*** Understand the business opportunities (including entrepreneurship) engendered by the big data phenomenon<br>***LO6:*** Gain insights about what working 'with big data' means through real life examples | • Feedback – Creation of a big data startup (through the school's incubator)<br>• Demo of the social network analysis tool<br>• Feedback from various projects | CEO of big data startup (current student of the school) |
| **4** | **The reality of big data for analytics experts** | ***LO7:*** Know the key players and dynamics of the analytics sector<br>***LO8:*** Perceive how analytics solutions can be used by organizations entering the data-driven era | • The big data landscape for firms specialized in analytics<br>• Big data solutions provided by the company<br>• Feedback from various projects | Director Business Solutions & Emerging Markets / large multinational company |
| **5** | **Technological impact of big data / Big data architecture** | ***LO9:*** Understanding the main big data architecture concepts | • Distributed databases / Hadoop<br>• What are the main technological impacts of big data in firms?<br>• Feedback – Projects in the aerospace and defense sectors | Managing enterprise architect and big data expert (consulting firm, big data leader in France) + Account manager (American software company, big data solutions) |
| **6** | **Big data and smart cities** | ***LO10:*** Understand how the public sector can benefit from the big data phenomenon<br>***LO11:*** Gain insights about the specificities/complexity of big data project management | • Promises and challenges of smart city projects<br>• Presentation of the smart city project lifecyle from the provider and the customer's points of view. | CEO of big data startup + digital manager of the city of Toulouse |
| **7** | **Big data, personal privacy, and intellectual property** | ***LO12:*** Reflect on the societal implications of the data-driven era.<br>***LO13:*** Understand the main ethical and legal issues raised by the big data phenomenon | • Big data ownership<br>• Legal aspects surrounding big data<br>• Privacy and intellectual property in the big data world | Professor of business law (expert in privacy, intellectual property, and Internet law) |
| **8** | **Big data strategy: Between jobs, processes, and collective intelligence** | ***LO14:*** Reflect on the new business models engendered by the big data phenomenon<br>***LO15:*** Reflect on the organizational changes occurring within organizations that decide to engage the path of data transformation | • Big data strategy and business models<br>• What are the main organizational impacts of big data in firms?<br>• Feedback from various projects (how the two startups collaborate) | Head of operations + CEO of two collaborating big data startups |

## Serious Game Context

The context of the serious game was a French newspaper, specialized in financial and economic news (and leader in the French market). Students were presented the main characteristics and difficulties of the newspaper industry such as the collapse of paper-based newspaper sales, the intense competition with web-based newspapers, and the overall tendency for customers to prefer digital information channels. The participants were presented the two main business objectives of the serious game:

1. Develop the usage of digital information channels (even free of charge) to expand online advertising revenues
2. Increase revenues in terms of memberships (two options), article readership, and content downloads.

To do so, participants had to: first get a good understanding of the dataset using a data visualization tool, identify patterns and potential opportunities, and finally derive business recommendations that are adapted to the context of the company and its industry. Students were not provided any guidance for any the three steps. The rationale behind this approach was to sensitize participants to the insertion of a data component into organizations' decision processes. The task was specifically designed so that students would follow the decision cycle: *business problem/opportunity* ⇨ *data* ⇨ *insights* ⇨ *business decision*.

The learning objectives of the serious game were the following:

- **LO16:** Understand what 'raw data' is about and how business value is in embedded into it.
- **LO17:** Be able to address a real-life business problem using a data-driven approach (by following the decision cycle: *business problem/opportunity* ⇨ *data* ⇨ *insights* ⇨ *business decision*.
- **LO18:** Be able to derive business insights from raw data using visual statistical techniques.
- **LO19:** Be able to efficiently communicate to an audience of experts in a big data context (mastery of big data terms, jargon, and key notions).
- **LO20:** Gain an understanding about how the job/function of managers is impacted in data-driven organizations.

## Dataset and Visualization Tool

The dataset was built on semi-real structured data collected from two leading French newspapers on which slight corrections were performed in order to ensure the pedagogical value of the dataset (to make sure that patterns could be identified by students). The data were stored on a simulated Hadoop cluster hosted on an external cloud platform. The database consisted of data from about 3.5 million newspaper users that had been captured for four consecutive months (providing a total of 14 million lines). About a hundred variables were provided to students including:

- **User Characteristics:** Such as gender, location, address, postcode, email address…
- **Membership Characteristics:** Such as membership status in relation to the various newspapers of the associated brand, membership duration, inactivity duration and degree…
- **Behavioral Characteristics:** Website connection numbers, number of received emailing campaign messages, number of downloaded pdf files…

The students were provided a full-day training to master the big data visualization tool (SAS® Visual Analytics) prior to partake the serious game. The demos and tutorial sessions used the dataset of the serious game to make sure that students had the data properly loaded.

The data visualization tool was provided by a leader in analytics solutions. Through a drag-and-drop graphical user interface, it allowed to explore datasets by identifying trends, patterns, and relationships. Basic functions included the manipulation of data such as creating hierarchies, aggregated measures, or joining datasets. Simple visualization functions involved standard descriptive statistical techniques such as bar charts, line charts, box plots, but also heat maps, correlation matrices, and word clouds. Predictive techniques could also be used such as the generation of band plots (based on the notion of confidence intervals) or decision trees (regression models) for instance. Figure 1 presents some of the graphs that can be generated when using the data visualization tool.

*Figure 1. Sample of visualization techniques provided by SAS® Visual Analytics*

### Deliverables and Presentations

The participants were provided four full days to prepare a 25-minute presentation of their main results and recommendations. The tool had two main usage modes: exploration and reporting. Students were given the choice to use the reporting functionality of the data visualization tool for their presentation (using a tablet as a remote) or else more conventional presentation software/ online applications. Four big data experts (the VP innovation of a consulting firm, a big data entrepreneur, a big data project manager, and an information systems professor specialized in big data) took part in evaluating the performance of each group.

An array of visual descriptive statistical techniques were used by participants to explore the various characteristics of the dataset, such as histograms, bar charts, pie charts, or bubble charts (see Figure 2 for sample graphs). More advanced uses included correlation tables, linear regressions, or confidence intervals (see Figure 3). The members of the jury acknowledged that most groups had gained a good understanding of the business case and an in-depth grasp of the dataset and its main patterns: e.g. four to five distinct readership profiles, the relationship between brand proximity and readership type/quantity, as well as factors correlated to revenues (direct or indirect).

## LESSONS LEARNT AND BIG DATA INSIGHTS

Following the two-week workshop, debrief sessions were organized with individuals from the business school as well as from the corporate partners that had been involved in the event and its organization. Direct feedback was provided by the participants during the two-week period while feedback forms were completed by all participants and collected at the end of the event. Participant observations were performed by the main organizers while coaching the students during the serious game, field notes were taken during the workshop, and informal interviews were conducted. The collected data allowed to derive insights for the organization of future big data events, courses, and programs. Lessons for organizations that are working towards nurturing a big data culture could also be identified.

### Collaborating with the Big Data Ecosystem

Big data is inherently complex. This results in an evolving and interconnected network of actors that interact (and often collaborate) with each other, covering a very wide spectrum of specialization domains such as applications (vertical, log data, ad/media), business intelligence, analytics, visualization, data/ infrastructure as a service, analytics infrastructure, and even traditional structured database specialists. As a result, big data projects usually require the expertise and services from several companies belonging to the overall big data ecosystem that includes all big data actors (including the many big data startups), companies specialized in big data integration and services (usually provided by consulting firms), and open source software projects such as Hadoop, Apache HBase, or MongoDB to name but a few.

From a pedagogical perspective, teaching big data is thus challenging. The complexity and many applications of big data can only be apprehended through direct involvement with the big data ecosystem. In simple terms, big data training programs, courses, and curricula shall be organized in such a way that students shall interact with an array of big data specialists in order to provide them a broad enough picture of the big data landscape. Furthermore, the big data ecosystem evolves quite fast in terms of both

*Figure 2. Sample of visual descriptive statistics techniques*

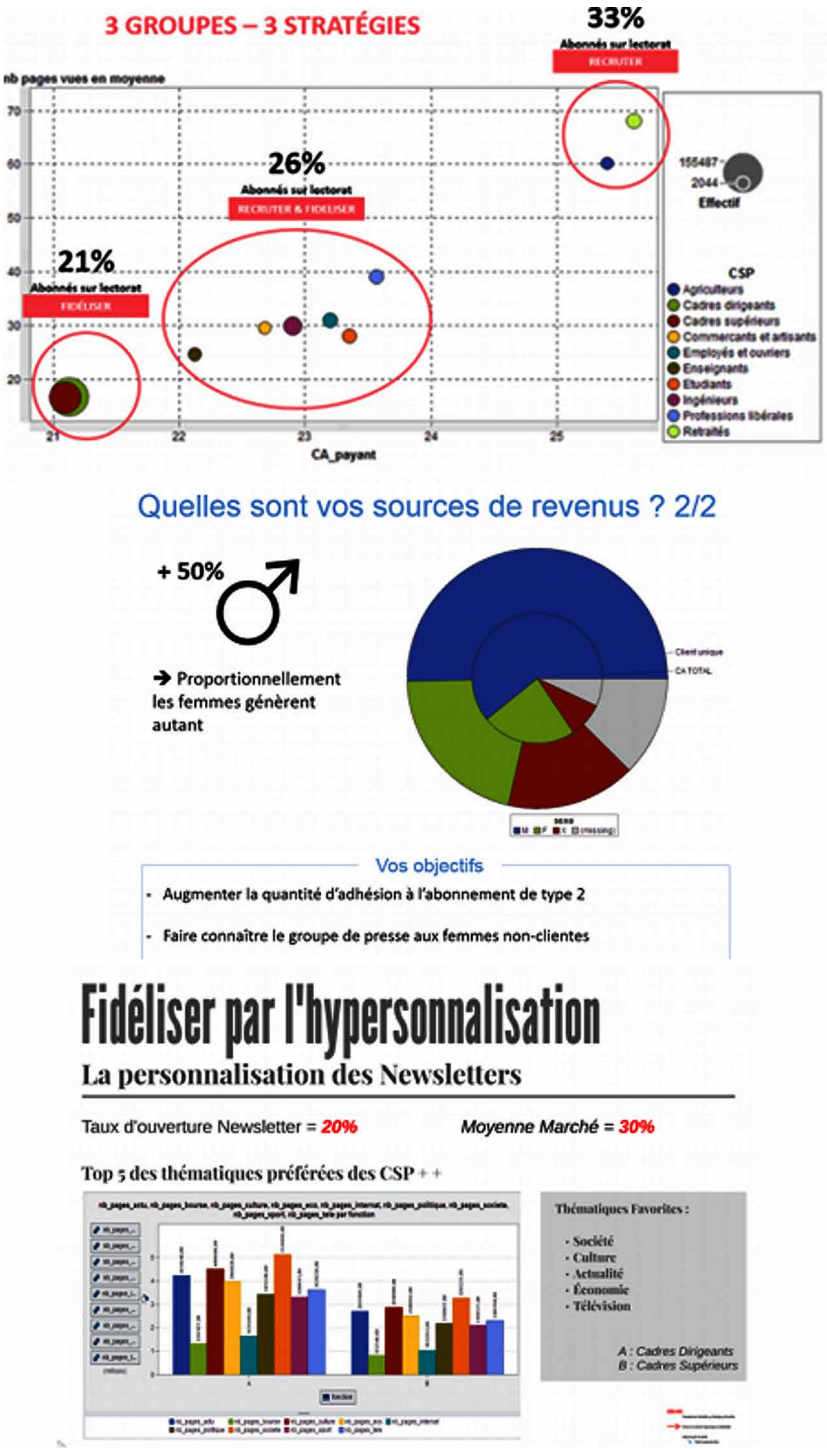
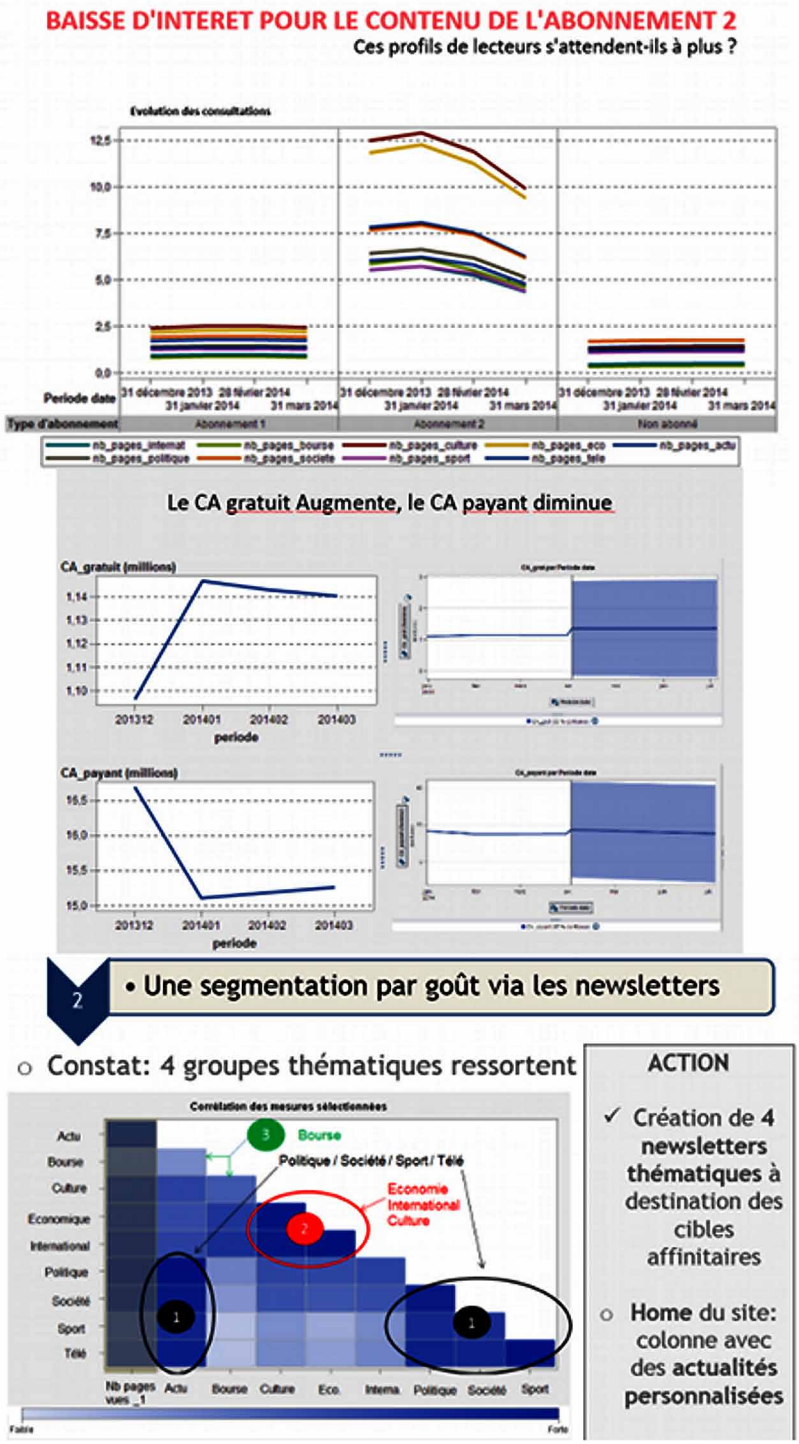


*Figure 3. Sample of advanced visual statistics techniques*



technology and actors. For instance, until the end of October 2013, Hadoop was seen as the breakthrough technology of the big data phenomenon, allowing to manage and transform large quantities of data with performance levels that had never been reached before. In 2014, Apache Spark dethroned Hadoop by allowing to work 100 times faster than Hadoop 1.0, and the trend will continue... The big data sphere is a bubbling ecosystem in which startups keep emerging and growing, and in which the sub-domain leaders are a mix of technology dinosaurs to mere infants with only a few years of existence.

As a consequence, higher education institutions are facing a nearly impossible challenge that is equivalent to being urged to catch a very important train that is already moving and which speed keeps increasing. The only way to tackle such challenge is to close the gap between educational institutions and the world of practitioners and engage into an active collaboration. Teaching big data cannot be efficiently done without being directly 'plugged' into the big data ecosystem. In other words, it is not sufficient to work towards reducing the dramatic big data skill shortage that is foreseen for the next two to three years. Big data is evolving at such a fast pace that new expertise, skills, and technologies keep emerging. It will not be astounding to hear from industry experts about another dramatic big data 2.0 skill shortage within the next few years. In short, the solution is to catch the train and stay in it while its pace increases instead of going down at the next stop and catching another train that goes even faster and is thus harder to catch.

## Touching Data to Feel its Business Value

Manipulating data is not a common task for managers who are more used to already digested and formatted information. It was interesting to notice that about a third of the time spent by all participants (the 20 groups) during the serious game was to get acquainted with the dataset and looking at raw data in order to get insights about the type, accuracy, and potential value of each variable. Participants used simple descriptive statistics for each of the variables but more importantly, they dedicated a lot of time displaying raw data and scrolling through the different values associated with each variable. Because there was no instructed sequence of tasks to be performed during the serious game (only the main objectives were provided), it can be argued with a reasonable amount of confidence that this common tendency about the need to 'touch' data naturally occurred. The participants quickly realized that it was simply impossible to derive relevant business recommendations without pulling up their sleeves, going back to data in its raw form, and spending time getting a good grasp of it.

By doing so participants were directly confronted with issues that were new to them but crucial to perform the task such as: dealing with missing variables and understanding the associated meaning, realizing that data sources are more or less reliable, understanding that the degree of accuracy of data sources can vary, realizing that sources can be trusted but not others… Participants gradually entered the world of data scientists and understood that deriving insights and making business decisions based on big data is no easy task. We noticed that the groups which dedicated the largest amount of time to get a good grasp of the dataset were the ones that performed the best (based on the performance scores provided by the members of the jury).

This validated the pertinence of the task that was designed for the serious game which was aiming at sensitizing future managers to the intrusion of data in the traditional decision lifecycle. If one has to define in simple terms what the main mission of business schools is when teaching big data, it can be argued that it is simply to create and consolidate the link between data and business decision in students'

conceptual mind maps. The coaching of students during the serious game allowed to clearly observe the gradual understanding from the participants that data contain business value and shall thus be treated as precious resources.

To answer the main question formulated in this chapter title, a sine qua non condition to engrain a big data mindset into our managers' DNA is to get managers acquainted to data and make them apprehend the business value that is embedded into data. This can only be done by touching data, manipulating them, and 'playing' with them.

## Playing with Data: Serious Games and Simulations

The big data business paradigm will engender a cultural shift in organizations. Whereas failure is used to be severely condemned in our nowadays highly competitive world, it becomes a positive mechanism in the big data world. Indeed, a big data mindset is characterized by an experiential culture in which one learns from failures and gradually improves through several cycles of experiments and actions. For instance, sentiment analysis modelling is an ever evolving task as language keeps changing: new words keep appearing (such as selfie…) and our society keeps giving birth to new ways of expression. For instance, more than 100 distinct emoticons are officially recognized (to which correspond a combination of two to three characters) and used in digital communication channels to express a broad range of emotions. Developing social media data models is thus an incremental process during which human actions and corrections allow to refine models up to satisfactory levels of accuracy and performance. In short, big data is not about success but rather about experimental failures that eventually lead to success.

Engendering an experiential culture within the mind of managers is a difficult task as it contradicts the basic nature of a manager's job. Time is money and failing costs even more money. To convince (future) managers of the benefits of experimentation in the big data context, the use of serious games or other types of simulation shall be recommended. Indeed, it is only through direct experimentation that one can seize that incremental learning allows to reach higher ends (at least in the big data context). To plant such seed in the mind of managers, well-designed big data simulations shall accompany participants through several loops of a fail-and-learn process and demonstrate the overall benefits of the approach.

The choice to use a data visualization task for a short big data training program (a few weeks) was highly satisfactory and shall be recommended for similar big data training initiatives. Indeed, a one-day training session was sufficient to ensure that participants had enough skills to perform basic but also advanced manipulations on the dataset. The technical barriers of entry were lowered as the drag-and-drop interface of data visualization applications are designed in such a way that users can go straight into the data without having to deal with either database management commands or obscure statistical functions.

Working in close collaboration with the big data ecosystem is crucial in the design of pedagogically effective simulations as their success is directly dependent on the provision of appropriate big data tools but also, and more importantly, on the use of datasets that contain a decent amount of 'pedagogical value'. Indeed, past and current experiences in designing big data events in the business school made us realize that good datasets are rare resources. Big data professionals are keen to provide real datasets (under certain terms) but the inherent complexity of big data renders the task to find quality datasets perhaps the most delicate one. In the case of the workshop presented in this book chapter, the opted solution was to rely on semi-real data that consisted of merging two real datasets (from two French newspapers) and to slightly inflate the patterns that characterized the final dataset (by applying several mathematical formulas on some of the columns and adding some random variation).

## Strengthening Statistics and Database Skills

Managers will in no way become data scientists. They will rather be an interface between a team of big data experts (data scientists, data analysts, big data solution architect…) and higher management. They may also be directly involved in big data projects within their organization. As a result, it is necessary that all participating actors shall use a common language and have some shared understanding on a number of basic big data-related concepts. In addition to getting big data engrained into their mindset, managers shall thus develop solid skills in statistics and databases. The workshop allowed to clearly identify the need for the two types of knowledge. The participants that had no prior knowledge on database management and data structure struggled throughout the serious game. Moreover, those that had not acquired a solid basis in business statistics were equally penalized.

Data visualization tools usually use statistical techniques that rely on concepts (such as descriptive statistics, linear regressions or time series…) that are covered in traditional business statistics courses provided by business schools. For instance, correlation is a common concept in a wealth of big data projects. Real-time decision making tools use simple and more advanced correlation techniques to identify insightful patterns that would otherwise remain hidden in exabytes or zettabytes of messy data. Online recommendation systems used by Netflix (Bollier & Firestone, 2010) or Amazon (Linden et al., 2003) look for correlations in users' viewing habits and product features. Making predictions is also an essential big data component and linear regressions (but also confidence intervals) are commonly used. As a result, it is crucial that business schools at least maintain the content of their conventional business statistics course and dedicate extra time and efforts to ensure that the main concepts are thoroughly acquired by students. This is a necessary condition to make sure that future managers will have a sound basis to evolve in a big data environment.

However, big data modelling techniques tend to rely more and more on advanced statistical concepts that are not taught in business schools. Extra efforts shall thus be spent to include them in curricula leading to globally stress out the importance of statistics in the training of future managers.

For example, predictive modeling techniques provided in data visualization tools often rely on generalized linear modeling, logistic regression, or classification trees, as it was the case with the tool used during the workshop. No single participant used such techniques to analyze the data. Cluster analysis and modeling is for instance widely used in online recommendation systems (Linden et al., 2003) but also to identify communities or commonalities within groups of users in social networks, or to identify customer segments to offer finely targeted services/products. Big data can help marketing research in performing simulations (e.g. Monte Carlo) to estimate entry market share for a new product/service, forecast sales, or evaluate the actual market size for a product or service. Finally, big data solutions heavily rely on correlation techniques to identify patterns among sources of data. However, big data experts are raising concern about the validity of such approach as correlation is far from equating causation which is what is usually sought. People such as Chris Anderson, editor-in-chief of Wired magazine claimed that big data had rendered obsolete the scientific method arguing that the analysis of correlation in large datasets would replace it. Big data experts have on the contrary shot an alarm signal at the overuse of correlation techniques and the associated theory-free causal inferences that are made. Spurious correlations abound in our world. A wrong conclusion about the existence of causation can lead to erroneous actions and negative consequences for businesses with ill-used big data tools. Expertise in the design of experiments as well as scientific models can help close the gap between causation and correlation (McAfee et al., 2012) and would be of great help for managers working on big data.

The Google Flu Trend case is one among the many examples of the limits of correlation. Google Flu Trends is a web-based service provided by Google that started in 2008. For several years, the theory-free service accurately predicted influenza outbreaks by correlating search engine terms and the spread of flu. In 2013, confidence in the accuracy of the service was strongly severed as it had drastically overestimated peak flu levels (Lazar et al., 2014). One of the potential explanations was that the media broadcasted a lot of negative 'flu' stories that provoked an increase of 'flu' web searches even by people who were healthy. A pharmaceutical company which would take for granted the predictions of the Google Flu Trend service and act accordingly would have lost a significant amount of money.

A similar argument can be made about the need to develop data structure and database management skills (which are not usually taught as a core course in business schools). This was a striking observation while coaching students during the serious game. For instance, the exploration of multidimensional data usually requires the organization of data into hierarchies allowing to slice and dice data through the use of filters. The participants that did not properly used data hierarchies (or did not use them all) remained at a superficial degree of comprehension of the dataset and provided shallow business recommendations. Difficulties in making the distinction between nominal, ordinal and interval variables, as well as the applicability of the different statistical techniques for each of the variable category, also appeared as an obstacle to perform good analyses. Finally, basic database management knowledge was a plus for the participants that were comfortable with the notions of database, tables, records, and fields, and who had previous experience in relational database management (such as handling SQL queries). The data visualization tool that was used during the workshop had a data builder mode which main functionality was to merge source tables from third-party vendor databases and manage the tables of the entire dataset. Participants who used the data builder module when exploring the dataset and understanding the link between the different tables, provided a much more in-depth analysis that resulted in quality business recommendations.

In conclusion, a necessary condition for business schools that are striving to diffuse big data into their curricula is to adjust the content of their business statistics courses and to emphasize the crucial importance of statistics in the big data context. Key database concepts (including managing and manipulating relational databases) shall also be taught as this provides a sound basis to understand important big data notions such as scalable distributed database systems or NoSQL.

## CONCLUSION

This book chapter aims at providing insights to business schools as well as organizations from both the private and the public sector, that are reflecting on the extent to which big data will impact the function and role of managers. Perceived as one of the most fertile terrains for taking advantage from the data transformation, the public sector is particularly targeted in this chapter.

By going beyond the conventional techno-centric view of big data that sees the phenomenon as simply a question of processing too large and too complex datasets, this book chapter adopts a higher-level societal approach that heralds a business paradigm shift in which data will play a predominant role and be a key business driver. A French top-ten business school reflected on how to address the overall big data talent shortage announced by most industry experts, who are also predicting that such skill shortage will also concern millions of managers throughout the world. The school designed a two-week big data workshop with the intention of experimenting a small-scale big data initiative (with 100 graduating

students for a master in management program) that would help gain insights for larger-scale initiatives aiming at diffusing big data into the various programs and curricula of the school. In addition to being an overall success by being perceived quite enthusiastically by all participants including the involved practitioners and the organizers, important lessons could be learnt. First, the nature of the big data phenomenon as well as the complexity of the big data landscape render the tight collaboration with the big data ecosystem a key success factor. Second, engraining a big data mindset into our managers' DNA can be effectively done only through the design of courses or programs that make students and trainees touch and manipulate data. Third, serious games and other types of simulation are ideal pedagogical strategies to infuse an experiential culture within the minds of trainees and to demonstrate the link that exists between data, business value, and business decision. Finally, skills in simple and advanced business statistics, but also data structure and database management appear to be crucial in establishing a conceptual and practical basis on which to build a sound understanding of big data in dedicated courses and programs. Considering the degree of commonality between French business schools and other international business schools, we believe that the derived insights and conclusions are also beneficial to any international business school that is at the reflection stage about how curricula shall be impacted by the advent of the data-driven era. We sincerely hope the overall approach as well as the pedagogical engineering presented in this chapter will provide insights to educational institutions and organizations from both the private and the public sector that have launched a similar reflection and have engaged on the path of the big data era.

## REFERENCES

Big Data. (n.d.a). In *Merriam-Webster dictionary*. Retrieved October 27, 2014, from http://www.merriam-webster.com/dictionary/big_data

Big Data. (n.d.b). In *Oxford dictionaries*. Retrieved October 27, 2014, from http://www.oxforddictionaries.com/definition/english/big-data

Big Data. (n.d.c). In *Gartner IT Glossary*. Retrieved October 27, 2014, from http://www.gartner.com/it-glossary/big-data

Bollier, D., & Firestone, C. M. (2010). *The promise and peril of big data*. Washington, DC: Aspen Institute, Communications and Society Program.

Brown, B., Chui, M., & Manyika, J. (2011). Are you ready for the era of 'big data'? *The McKinsey Quarterly*, *4*, 24–35.

Chen, H. (2011). Design Science, Grand Challenges, and Societal Impacts. *ACM Transactions on Management Information Systems*, *2*(1), 1–10. doi:10.1145/1929916.1929917

Chen, H., Chiang, R., & Storey, V. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *Management Information Systems Quarterly*, *36*(4), 1165–1188.

Galliers, R. (2003). Change as crisis or growth? Toward a trans-disciplinary view of information systems as a field of study: A response to Benbasat and Zmud's call for returning to the IT. *Journal of the Association for Information Systems*, *4*(6), 337–351.

Gartner, Inc. (2012a). *Gartner Says Big Data Will Drive $28 Billion of IT Spending in 2012*. Retrieved October 27, 2014, from http://www.gartner.com/newsroom/id/2200815

Gartner, Inc. (2012b). *Gartner Says Big Data Creates Big Jobs: 4.4 Million IT Jobs Globally to Support Big Data By 2015*. Retrieved October 27, 2014, from http://www.gartner.com/newsroom/id/2207915

Gil Press. (2013). *$16.1 Billion Big Data market: 2014 Predictions From IDS And IIA*. Retrieved from http://www.forbes.com/sites/gilpress/2013/12/12/16-1-billion-big-data-market-2014-predictions-from-idc-and-iia

Lazer, D., Kennedy, R., & Vespignani, A. (2014). The Parable of Google Flu: Traps in Big Data Analysis. *Science*, *343*(6176), 1203–1205. doi:10.1126/science.1248506 PMID:24626916

Linden, G., Smith, B., & York, J. (2003). Amazon.com recommendations: Item-to-item collaborative filtering. *Internet Computing*, *7*(1), 76–80. doi:10.1109/MIC.2003.1167344

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big data: The next frontier for innovation, competition, and productivity*. Retrieved October 27, 2014, from http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big Data. The management revolution. *Harvard Business Review*, *90*(10), 61–67. PMID:23074865

Olofson, C. W., & Vesset, D. (2012). *Big Data: Trends, Strategies and SAP Technology*. Report ICD #236135.

Schoenherr, T., & Speier-Pero, C. (2015). Data Science, Predictive Analytics, and Big Data in Supply Chain Management: Current State and Future Potential. *Journal of Business Logistics*, *36*(1), 120–132. doi:10.1111/jbl.12082

Sondergaard, P. (2013). *Gartner Analyst Opening Keynote Gartner. Symposium/ITxpo 2013*. Retrieved October 27, 2014, from http://www.gartnereventsondemand.com/index.php?t=trailer&e=SYM23&i=K2

Ward, J. S., & Barker, A. (2013). *Undefined By Data: A Survey of Big Data Definitions*. arXiv preprint arXiv:1309.5821

Wixom, B., Thlini, A., Douglas, D., Goul, M., Gupta, B., Iyer, L., & Turetken, O. et al. (2014). The Current State of Business Intelligence in Academia: The Arrival of Big Data. *Communications of the Association for Information Systems*, *34*(1).

## KEY TERMS AND DEFINITIONS

**(Apache) Hadoop:** An open source project that provides a programming framework allowing the scalable and distributed processing of large datasets across a cluster of commodity servers. It is part of the Apache project and sponsored by the Apache Software Foundation.

**(Apache) HBase:** A column-oriented database management system that is used to provide real-time read and write access to datasets hosted on Apache Hadoop clusters. Apache HBase does not relies on the support a structured query language such as SQL. It is a sub-project of the Apache Hadoop project.

**(Apache) Hive:** An open source data warehouse system designed to query and analyze large datasets stored in Hadoop files.

**(Apache) Spark:** An open-source parallel processing framework that allows developers to run large-scale data analytics applications across a cluster of commodity servers. The Spark project fits within the Hadoop community.

**(Big) Data Visualization:** The implementation of traditional and more contemporary visualization techniques that help identify trends, patterns, and relationships within data.

**Machine Learning:** A scientific/engineering field that aims at constructing algorithms that can learn from data and at building models that are typically used to make decisions or predictions.

**MapReduce:** A software framework used at the core of the Hadoop system and allowing developers to write programs for processing large amounts of unstructured data in parallel across a cluster of commodity servers.

**MongoDB:** An open source project that provides a cross-platform document-oriented database framework. It is among the most popular NoSQL (see definition) database systems.

**NoSQL:** (also called Not Only SQL)**:** An alternative database environment (as opposed to more conventional relational databases) that allows to design and manage very large sets of distributed data.

**Python:** An open-source high-level programming language (object-oriented) that is known for being easy to read and simple to implement. It is often used in big data projects.