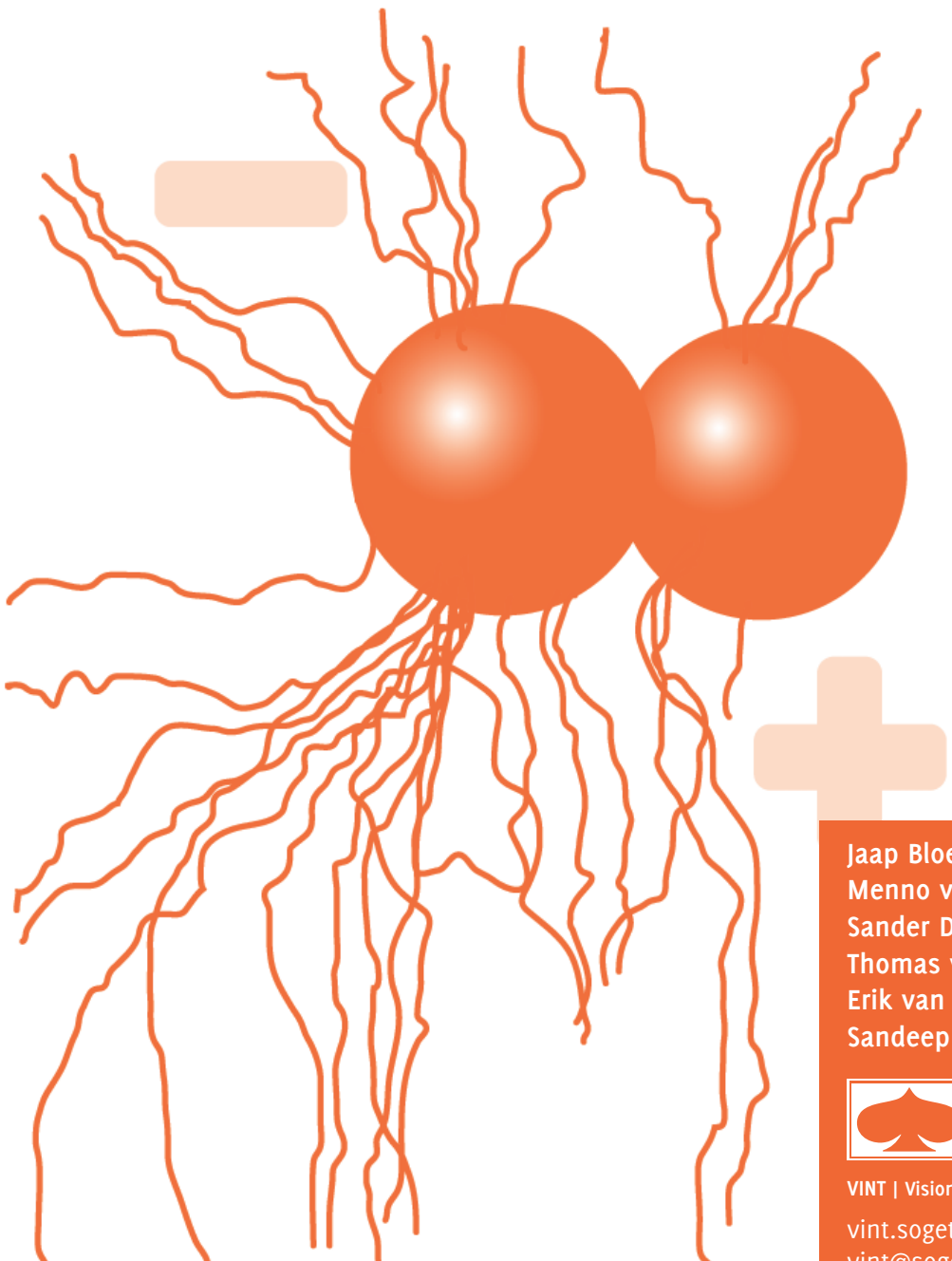
VINT research report ① of 4
VINT research report ② of 4
VINT research report ③ of 4
VINT research report ④ of 4

Your Big Data Potential

The Art of the Possible



Jaap Bloem
Menno van Doorn
Sander Duivestein
Thomas van Manen
Erik van Ommeren
Sandeep Sachdeva

SOGETI

VINT | Vision • Inspiration • Navigation • Trends

vint.sogeti.com
vint@sogeti.nl

Table of Contents

Book production LINE UP boek en media bv, Groningen

SOGETI

VINT | Vision • Inspiration • Navigation • Trends

VINT's Big Data Research Reports

Since 2005, when the notion of "Big Data" was coined – by O'Reilly Media remarkably enough, which had presented Web 2.0 a year previously – it has demanded increasing attention. In terms of technological development, the field of Big Data has been moving rapidly, and that is an understatement.

In *Creating Clarity with Big Data*, the first of our four research reports, we respond to the question of what Big Data actually is, where it differs from existing data-based references, and how the transformative potential of Big Data can be estimated.

The concrete adoption of new policy and plans in organizations form the main theme of our second report, entitled *Big Social*: *Predicting Behavior with Big Data*, which was inspired by the social network activity of Web 2.0 in particular.

The data explosion is occurring all around us, but an important part of the discussion concerns the extent to which organizations ought to commit themselves to Big Data. The answer is: they should do so on the basis of well-considered policy. Policy, whether it comes from external or internal sources, has to cope with the privacy issue, which is the subject of our third research report, *Privacy, Technology and the Law*.

With four Big Data reports, VINT aims to create clarity by putting experience and vision in perspective: independent and supported by examples. It is impossible to provide satisfactory and exhaustive answers to many of the issues that arise, and probably they raise even more questions – about strategic choices that you have to make, for example. We have reserved this fourth report for exactly that theme. This report is organized around ten simple questions that you need to consider.

We want to extend our discussion with you about this new data frontier, extend it online at http://vint.sogeti.com/bigdata and, of course, in personal exchanges. By actively participating in this discussion, you will help all of us hone our thinking about Big Data, and we can all arrive at clear and responsible decisions as a consequence of progressive insight.

"The Art of the Possible"

Intensive focus on business, organization and technology

In the nineteenth century, the German statesman Otto von Bismarck referred to politics as "the art of the possible." This also applies to Big Data: operating cautiously, while simultaneously pulling out all the stops with the aim of maximizing results, clarifying decision-making, and stimulating new insight. The seven conclusions and recommendations presented at the end of this report dovetail perfectly with this aim. Here is a brief summary:

Big Data is the new, intensive, organization-wide focus on business, organization and technology. Accordingly, you must ensure that the organization has sufficient technological and analytical expertise, as well as the appropriate digital and organizational competences. After all, your aim is also to excel digitally and operationally. This is possible because making the best of Big Data in a structural way is becoming increasingly affordable.
With regard to Business Intelligence, *Data Discovery* is the next phase. It helps you combine lucid "magic moments" in your business operations with a significantly better performance through interactive visualization, exploration, planning and execution.

To start with, you can fire your imagination in inspiration sessions, followed by one or more concretization workshops in which you determine, in conjunction with an organization-wide team, where lucrative Big Data initiatives can be developed to suit your situation.

We analyze this constellation of potential in our final Big Data Research Report by distilling it into ten questions. These are presented in the Table of Contents, and we shall deal with them shortly, one by one. But first of all, we need to understand the importance of keeping an open mind.

Not science fiction but science facts

Charlie Beck, police commissioner in Los Angeles, is clear about the goal: there was no more money available, and no more police officers, so creativity simply had to increase. As the result of a political choice, all the stops had to be pulled out and results had to improve. For that reason the police force began to use a Predictive Policing algorithm, a Big Data solution that is now being applied in more than ten major American cities. In Los Angeles, crime decreased by 13 percent, and in Santa Cruz, which is also in California, by 26 percent.

The analytical software was developed by two mathematicians, an anthropologist and a criminologist, and is founded upon a model that predicts the aftershocks of earthquakes. For example, a criminal often returns to the scene of a previously committed crime. Such *aftercrimes* follow the same pattern as the aftershocks of an earthquake.

On the basis of location, time, and type of crime, the software is capable of defining "prediction boxes" of 500 square meters.

This resembles science fiction, but is actually science fact. Historical facts and real-time events contain crosslinks and correlations of which we were unaware until Big Data came along. The retail trade works in a similar way. The case of the Target supermarket chain is iconic in that respect, continually growing its sales by perfecting the art of targeting. For example, data scientist Andrew Pole was able to predict whether or not a consumer was pregnant and the likely date of birth, on the basis of product purchases. This kind of predictive skill lies behind the Big Data potential that everyone is talking about these days. Take the Fraunhofer IAIS, for instance, the German institute for Intelligent Analysis and Information Systems. After extensive investigation on the basis of desk research, expert workshops and a survey, this Institute presented the Big Data innovation potential for German enterprises in March 2013 (see Question 2).

Big Data is your new colleague

Capital One Labs, a part of the Capital One bank and credit card company, is one of the many organizations that recruit Big Data talent. On Kaggle, a platform where many data professionals convene, we observed a Capital One job advertisement for this kind of data scientist. Anyone who applies for a job with Capital One Labs knows that he/she may be entering a Silicon-Valley-like culture, and will be given the following extraordinary assignment:

> *"to push the envelope to explore The Art of the Possible"*

In order to tap into the potential of Big Data, people have to understand the art of exploring the possible and of making the apparently impossible possible. ING bank articulates similar thoughts in its quest for new Big Data innovators: you have to stimulate out-of-the-box thinking, while showing that you "always behave in that way yourself," as the ING advertisement on the Monsterboard job vacancy site says (see Question 4). Big Data is your new colleague and the mission is to import and implement radical ideas: science fiction based on science facts. For instance, Agentschap Telecom is intensively engaged in the development of new methods, including techniques for tracing pirate radio stations. Johannes Brouwer, the head of the IT department there, states:

> *"The only restriction in the domain of Big Data is one's own imaginative power."*

Spectacular results and promising experiments

Vigorous growth in turnover and spectacularly decreasing crime statistics – that is what the vast majority want to see. Of course, there are always restrictions – we shall mention many of them in the course of this report – but Big Data is genuinely "The Art of the Possible," and all new technologies make it possible to tackle matters in a radically different way.

Wal-Mart applies an intensive Big Data strategy by means of which it follows hundreds of millions of keywords and purchasing and searching behavior obtained via Google, Bing, Facebook, Groupon, Yelp, etc. These are imported via APIs after which the *bidding engine, analytics engine* and *pricing engine* are let loose upon them (see also walmartlabs.com/platform). Wal-Mart Labs, the "social data R&D" department, is seriously engaged in experimenting with, for example, semantic analyses. In this way, the so-called *Social Genome* is charted, consisting of detailed profiles of customers, topics, products, locations and events. The first results are now visible, such as the Wal-Mart app called Shoppycat, which presents gift ideas based on algorithms that interpret updates from social media. Presently, there are many such exciting new ideas and business cases.

Wait no longer: tap into your Big Data potential

After the first phase of Big Data projects and experiments, the players are now beginning to seek more intensive contact with one another. To help organizations develop their data intelligence, dataset suppliers and analysis partners are active at opposite ends of the sector. Of course, the major IT partners are also involved. You yourself must determine with whom you wish to establish business connections, as well as the extent to which you wish to retain control of your Big Data activities in the long run.

MyBuys supplies both data and analytical capacity. The personalization engine of this young company is based on more than 200 million customer profiles and 100 terabytes of data in order to deliver real-time recommendations. MyBuys currently has more than 400 client organizations who are seeking to improve their sales. To financial service providers such as Capital One, which performs more than 80,000 Big Data experiments a year, there are many more possibilities based on aggregated transaction information, as long as they lie within the legal framework. The message to entrepreneurs is clear: seek a reliable partner and simply begin, according to the Fraunhofer IAIS, because:

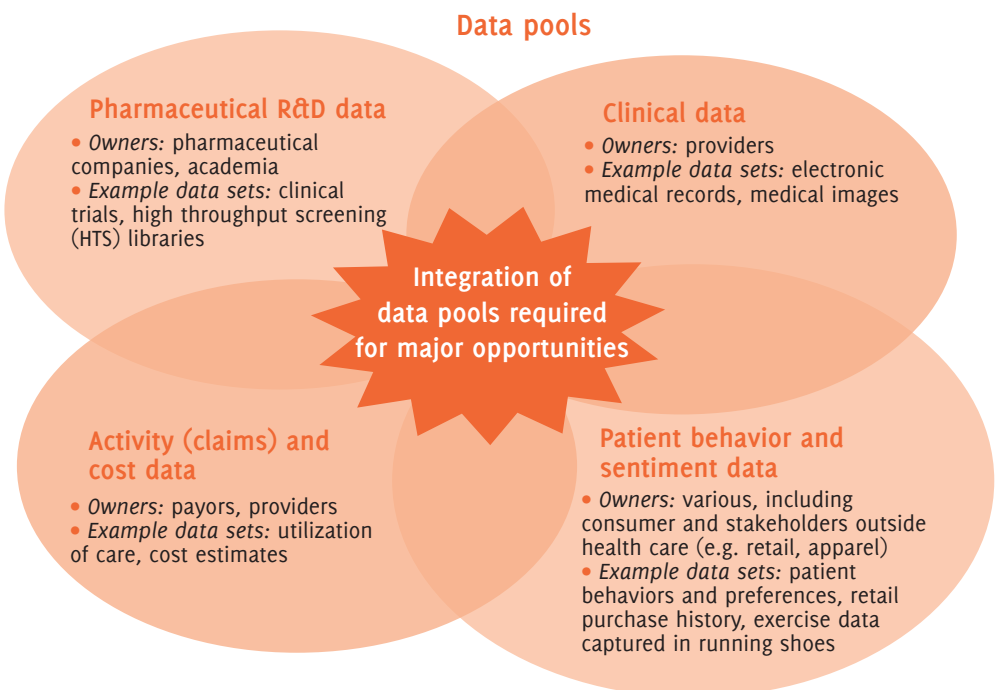
> *"ranging from sensor data to Business Intelligence, from media analysis to visual information system, you are capable of doing more with data."*

The core objective of all Big Data initiatives is to look beyond your own confines and to seek interesting combinations of internal and external, structured and unstructured data. This may be simply combining the current location of someone's telephone and transactions executed with his or her credit card. Too much distance between these may indicate fraud, certainly in combination with heightened transaction frequency. In this way, the data of telcos and banks could form a new kind of service. We do not explore all possible trans-sectorial connections in this report. Still, in concluding this introduction we need to emphasize that Big Data's potential can only be fulfilled when all intra- and inter-sectorial inefficiencies have vanished.

We shall take the healthcare sector as an example, but the energy or transport sector would be just as relevant. Healthcare provides many examples, such as Philips with its "hospital to home" strategy. Philips develops new products and services that give doctors, pharmacies, nursing personnel and even patients the opportunity to organize healthcare in a different way, with data technology and data visualization techniques.

Walgreens, the largest pharmacy in America, provides another example. It uses Big Data in linking point-of-sale data with social media, with data from customers' wearable computers, with data from partners such as the Nationwide Health Information Network, and with clinical data, all with the ultimate aim of improving patient care. Big Data potential radiates from the *data pools* that emerge from these kinds of new collaborations, as is shown by this overview of American healthcare presented by the McKinsey Global Institute:

Four distinct big data pools exist in the US health care domain today with little overlap in ownership and low integration.



Source: McKinsey Global Institute (2011)

The Big Data "Art of the Possible," based on mixing and matching, has demonstrably vast potential, but also requires much expertise, cooperation and coordination with regard to organization and levels of data and technology.

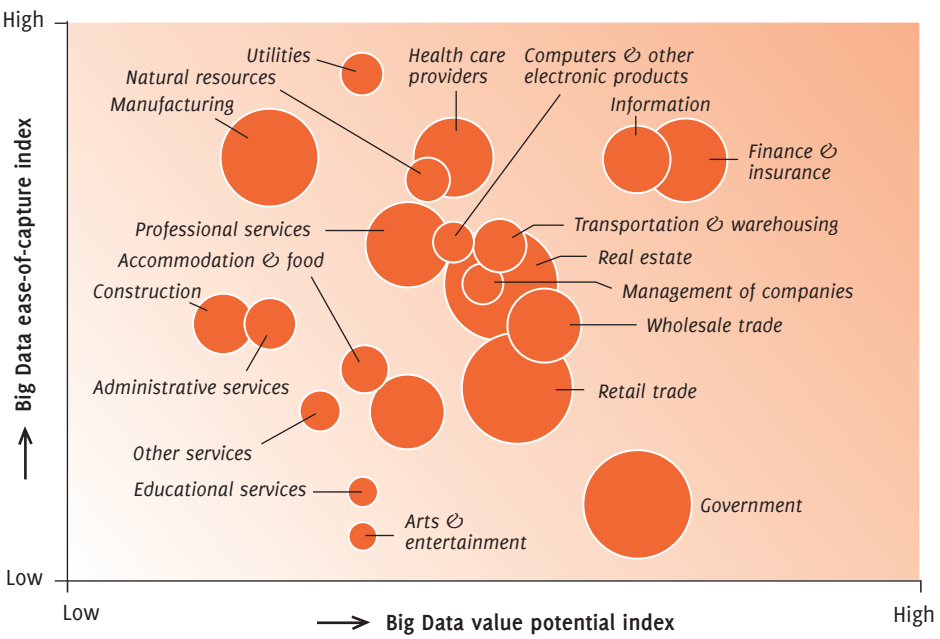
Your Potential: The Tension Between *Could Be* and *Is*

After the iconic Big Data bubble chart (*Value Potential* versus *Ease of Capturing*) by means of which the McKinsey Global Institute predicted, in 2012, hundreds of billions of dollars in yields for various sectors of the American economy, there followed a whole procession of case descriptions that heralded a new data-intensive era. Potential applications were served up, technology was available, and the first smartly applied Big Data initiatives were already proving their productivity. Organizations were convinced: the status of Big Data as *The next frontier for innovation, competition and productivity* – the report with which the McKinsey Global Institute kicked off the new trend in 2011 – seemed indisputable. Perhaps there were no best practices to follow at that time, but "emerging next practices" had already left their mark, as Michael Chui of McKinsey remarked at the MIT Sloan CIO Symposium precisely a year later.

The ease of capturing Big Data's value, and the magnitude of its potential, vary across sectors.

Example: US economy

Size of bubble indicates relative contribution to GDP



Source: McKinsey Center for Business Technology (2012)

We are now a year further along, and the next step is, of course, to consider how Big Data's potential is currently being realized and how far we have advanced in this direction. At the end of 2012 and the beginning of 2013, a number of interesting studies were performed on ways in which organizations were engaged in developing their Big Data potential. On the basis of these studies, the German Fraunhofer IAIS (the

Institute for Intelligent Analysis and Information Systems), TCS (Tata Consultancy Services) and SAS, in conjunction with the CMO (Chief Marketing Officer) Council, responded to the question about the potential of Big Data, approaching the question from several complementary angles.

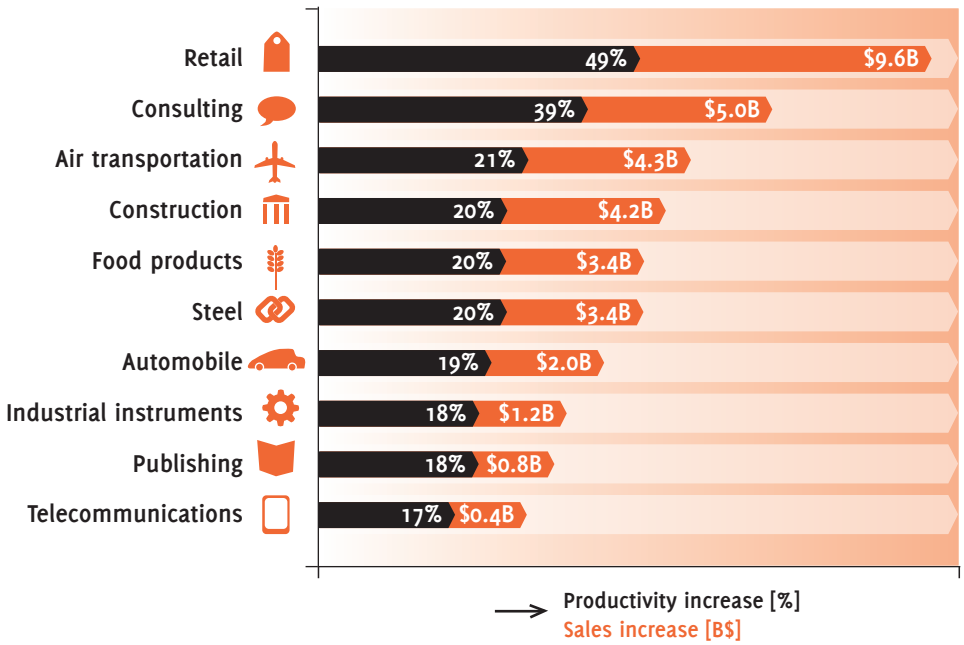
They give new insights into the way in which the ten simple questions (which are presented once more at the end of this section) can be answered. They thus chart the tension between *Could Be* and *Is* – thus, your potential. On every occasion, in order to make well-considered decisions, you must measure your own position and ambitions against your potential and the concrete lessons you've learned.

You should not abandon past data-related achievements. On the contrary, you can examine them more closely, broaden them, expand them and especially integrate them. But you should only begin this integration with a good knowledge of details – of developments in technological and organizational fields, drivers, needs, restrictions, requirements and impact. Act according to your best insight with a lucid priority plan, at your own speed and with clear aims and performance indicators. You may decide to disregard this kind of general advice in favor of addressing these details in a concrete way and, above all, by being honest about the situation throughout the organization and involving all stakeholders.

Accordingly, this fourth Big Data report by VINT is no simple roadmap but, following our Big Data reports 1, 2 and 3, it is a concrete and solid conclusion built around ten straightforward core questions. We envisage Big Data integration as the basis of new Business Intelligence; we see technology that enables a data-intensive approach to the issues touched upon in the questions; we examine your plan to become data-driven; we analyze your degree of Digital Advantage, and we conclude with a forecast of Big Data potential in 2020. Now is the time to develop your Big Data potential, and that is why you will find a checklist of twenty items at the end of this final Big Data research report.

Our knowledge and experience with what we call Big Data – immense amounts of data, and/or very varied data, and/or very rapidly changing data – is expanding day by day. Originally the notion of Big Data was linked to *Volume*, *Variety* and *Velocity*, but another three Vs have now joined up: *Veracity*, *Variability* and, of course, *Value*. The last three form the universal yardstick against which we must measure all data-intensive activities because, quoting Tom Davenport the analytics guru, we wish to use them to move from *descriptive* to *predictive* and ultimately *prescriptive analytics*. If the predictive analyses turn out to be right, we will win time and effectiveness by immediately prescribing and adhering to them.

Big Data is thus a wake-up call to use all conceivable data that we have at our disposal in order to become genuinely "data-driven." The following figure shows the *Value* component of effective data usage, presented for ten different business sectors.



Source: University of Texas (2011)

Isn't it marvelous? But how do you do it? The website http://UnlockingBigData.com offers a simple Big Data self-test, and this should supply all the necessary answers. Just answer the questions in the categories of *Data, People, Technology, Process* and *Intent*, and you will have a snapshot of how "mature" your business is, with epithets such as *Novice, Beginner, Competent, Proficient* and *Expert*. Good to hear, but you can probably already state with a reasonable degree of accuracy just where you are located on the Big Data scale.

In question 8, about what is looming on the horizon, we present the results of a much-renowned study by MIT in conjunction with Capgemini. The conclusion is that a digital strategy does pay off. Companies that invest more in digital and organizational effort have a higher turnover, more margin, and better market value. Big Data is the latest new development in the context of your digital strategy.

Removing ROI (Return on Investment) from your Big Data activities, closing the gap between potential and reality, begins with posing the right questions. Many questions have already been dealt with in our previous Big Data reports. You can consult them again to gain some inspiration. This final report lists the following ten most important questions in order to help formulate a concrete plan. We offer a number of key points, by means of which you can draw your own roadmap.

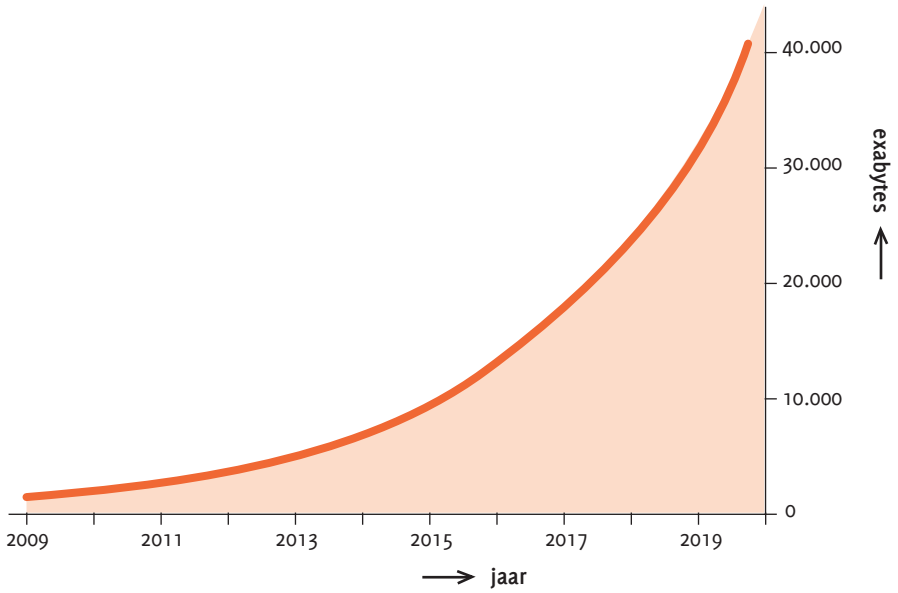
In technological terms, questions 6 and 7 pinpoint the difference with traditional RDBMS environments, but a purposeful data-focus on the combination of business, organization and technology is the core as well as the aim of the exercise.

Question 1: Why Big Data intelligence?

Answer: Because large amounts of data become available at little cost. These data contain valuable insights and their processing is attainable in technological and analytical terms.

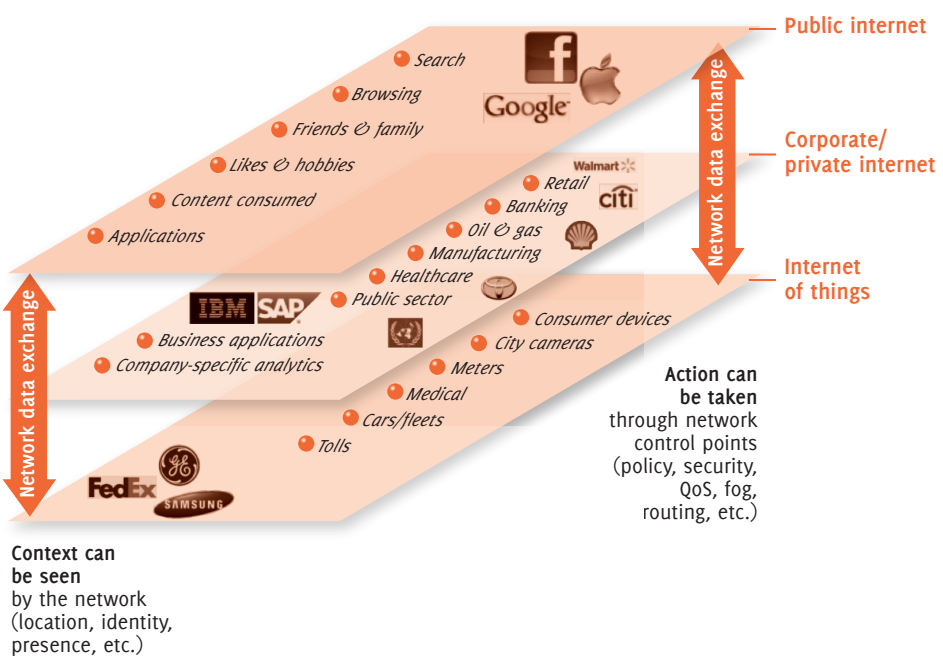
In 1965, Gordon Moore made a prediction in the *Electronics Magazine* that the number of transistors in a chip would double every two years. Until the present day, that prediction has been true every year, and it is now referred to as "Moore's Law." The two-year constant has now been adjusted to eighteen months. In the past thirty years, this exponential growth has become visible not only in the mathematical power of processors but also in working memory, storage space, bandwidth, the number of electronic sensors, and the quantity of data.

In its study entitled *The Digital Universe in 2020*, IDC calculated that the second decade of this century would witness a growth in the quantity of data to 40,000 exabytes, in other words: 40 billion gigabytes. This is equivalent to 5,200 gigabytes for each person on Earth. In 2005, the counter indicated only 130 exabytes.



In this decade, the amount of data worldwide will grow to 40,000 exabytes
Source: IDC & EMC (2012)

According to IDC, a quarter of the current Big Data mountain contains information that is useful for analysis. In 2020 this will grow to one third. Therefore we will be dealing with exabytes of valuable data in the coming years. In its working paper entitled *Defining a Taxonomy of Enterprise Data Growth*, the Center for Large Scale Data Systems concluded that most data that are currently being created within organizations are unstructured data. The data originate in a combination of three internet types: the Public Internet, the Internet of Organizations, and the Internet of Things.



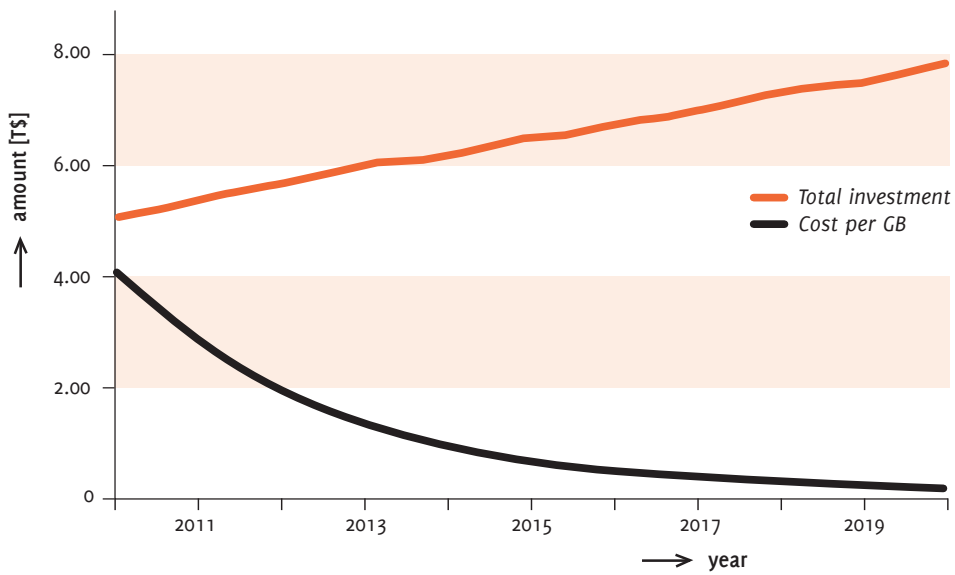
Source: Cisco IBSG (2012)

The current decade will largely be concerned with the question of how organizations will tap into this immense Big Data potential and what exactly that will mean. MIT professor Erik Brynjolfsson compares it to the invention of the microscope, which also enabled great surprises and new insights in many fields and disciplines:

> *"The microscope made it possible to see and measure things in a way that was unprecedented. Now we are dealing with the modern equivalent of the microscope."*

Thanks to advanced hardware and software, we are now capable of zooming in and out at great speed, with the aim of discovering structures and links that will give us more insight and enable us to make better decisions and find more effective solutions. Marketing, healthcare, energy supplies, transport, every type of service provision – in short, all forms of applied science take on a completely new aspect, and this will have an unprecedented impact.

This will also bring opportunities for the IT sector itself, which will also undergo transformation. Investments in IT will rise thanks to the need for Big Data insights, while the costs per gigabyte will decrease from four dollars to a few cents, primarily due to the explosive increase in data.

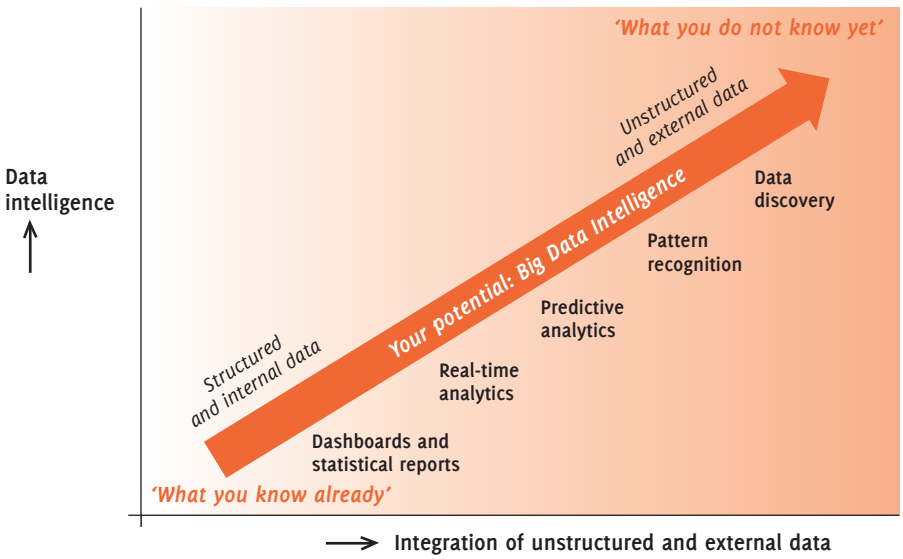


In 2020 one gigabyte will cost almost nothing and investments will rise substantially
Source: IDC & EMC (2012)

This picture is confirmed by the results of a study entitled *Big Data: Big Potential, Big Priority* that Cisco published at the end of March 2013. It presented the results of a survey on the status of Big Data in eighteen countries. More than half of the companies expected to make higher investments in the coming years thanks to the Big Data priorities that are now being specified.

Potential lies in unstructured data and business transformation

Capitalizing on Big Data potential goes hand-in-hand with investment aimed at getting hold of unstructured and external data: begin on a small scale and subsequently build up the capacity to harvest the unstructured and external data. That is the challenge, that is the Big Data potential that is consistently mentioned: accessing unstructured and external data to develop new insights. That is, concisely summarized, what the term "next-gen Business Intelligence" refers to. It is ultimately a question of "data intelligence." It represents a whole new generation of approaches, tools, insights, and different ways of working (faster, better and much more efficient).



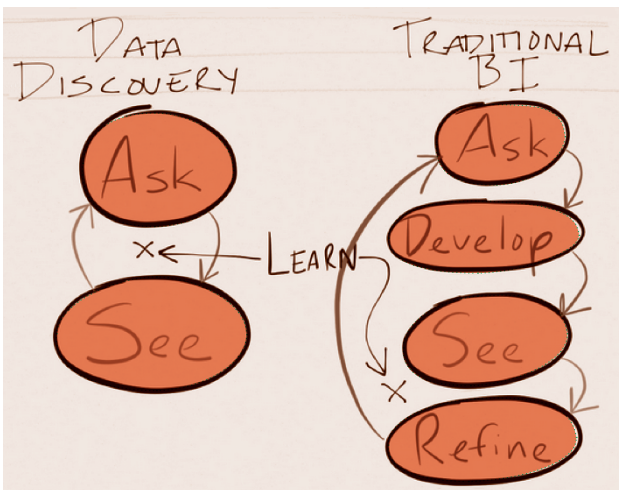
Business Intelligence on the basis of Big Data integration leads to new insights, and faster and better decision-making

In the meantime, the flood of success stories continues to grow. Shell detects oil fields through Big Data, the Netherlands Forensic Institute investigates data to trace criminal behavior, the Dutch Telecom Agency monitors social media to protect the airwaves, and KLM uses Big Data to gain better insight into the search behavior of customers. But everywhere on the globe, capitalization on and acceptance of Big Data are still in their infancy.

Nevertheless, data integration and Big Data are already familiar faces in some organizations and among individual experts, think tanks, and particularly in more scientific and data-intensive organizations such as forensic institutes, academic hospitals, telecom companies, banks, insurers, credit card companies and energy suppliers. In such settings, Big Data technology complements traditional research methods such as data-warehousing, data-mining and Business Intelligence. In several cases, in-house experiments are carried out with new technologies, and the impact of the new technology on the existing operational landscape is closely scrutinized.

Nowadays there is a much greater volume of data available from customers and competitors than was the case roughly five years ago. And much greater data volume is certainly coming. Fortunately, new technology, such as Hadoop and NoSQL ("Not Only SQL"), is already capable of coping with growth. This development brings many new opportunities, such as the upsurge of the so-called *Data Discovery*, which enables an interactive, visual, rapid-cycle question-and-answer analysis on the basis of modern data technology, enabling more rapid insight.

Capital One, the fifth largest credit card company in America, indicates that 80,000 data trials are performed annually. In this way, important new possibilities are discovered and the ROI of every marketing dollar is maximized. (See an up-to-date overview of Data Discovery tools on the website of Applied Data Labs.)



Data Discovery is the next-generation Business Intelligence

In terms of technology, the potential of Big Data lies in its capacity to interpret and apply external and unstructured data. In terms of organization, we are dealing with a major transformation. The Big Data potential of every organization is developed along these two axes – the capacity to process unstructured data and the capacity to transform your organization.

In its report, McKinsey explicitly accentuated the necessity of transformation. Without this transformation, there can be no mention of efficiency. This is much more than simple rhetoric. The expected leverage effect of Big Data is described by Viktor Mayer-Schönberger, professor at the Oxford Internet Institute, and Kenneth Cukier, Data Editor of *The Economist*, in their latest book:

> *"[Big Data] refers to the things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value in ways that change markets, organizations, the relationships between citizens and governments, and more."*

Cukier and Mayer-Schönberger's essay "The Rise of Big Data" was featured in the April/May 2013 edition of *Foreign Affairs*, the leading American magazine on international relations. This demonstrates just how topical and critical Big Data and its potential applications are, globally. This engagement is expected to continue into the foreseeable future.

Question 2: What new insights can I expect?

Answer: Two studies, one by Fraunhofer in Germany and the other by TCS, indicate where the greatest global potential of Big Data lies at present.

Big Data – Advantage through knowledge: Innovation potential analysis 2013

In March 2013, the German Fraunhofer Institute for Intelligent Analysis and Information Systems (IAIS) and the Ministry of Economic Affairs and Technology revealed how organizations in Germany, the fourth largest economy in the world, currently regard the innovation potential of Big Data. IAIS is a part of the Fraunhofer organization for applied research and, with more than 20,000 employees in Europe, it is the largest institute of its kind. The Institute categorizes the present use of Big Data as follows:

1. Monitoring markets to identify sales opportunities
2. Creating personalized offers
3. Recognizing repeat customers at an early stage (*churn*)
4. Recruiting staff
5. Forecasting sales, for planning and control
6. Predicting wear and maintenance
7. Directing management, decision-making & control
8. Detecting fraud
9. Estimating financial risks
10. Recognizing cyber attacks
11. Improving products
12. Developing innovative products

The *Innovation Potential Analysis 2013* formulated by Fraunhofer IAIS comprises three sections: an inventory based on theoretical research, five sectorial expert workshops, and an online survey. This last element produced the following seven main results:

1. In general, trade & industry are aware of the possibilities that Big Data analysis offers.
2. The greatest potential lies in building up a strategic competitive advantage (69 percent), followed by a higher turnover (61 percent) and cost-savings (55 percent). Higher productivity and data-driven planning and decision-making are also mentioned as objectives.
3. Not one but several organizational departments want to benefit from the advantages provided by Big Data, and a large group mentioned management in particular.
4. Marketing and operational aims score highly, in contrast to aims relating to IT.
5. At this moment, organizations are not optimally configured for Big Data analysis. The budget is too small, responsibilities and governance are not clearly specified, and competency needed to handle Big Data efficiently is inadequately developed.
6. Various issues will be dealt with in the coming five years, and no one anticipates a decline in the Big Data budget.
7. In order to be able to master the required competences, most respondents would like to learn about the best practices and are willing to undergo training.

The five sectorial expert workshops

Expert workshops were held for the following sectors: financial service provision, telecom and media, insurers, trade and retail, and market research. The aim was to generate a roadmap for the future. To start with, it is evident that Big Data is not exclusively an IT theme, and it does serve strategic aims. Moreover, it is also clear that interorganizational and intersectorial cross-fertilization, as well as a combination of findings and points of concern, may constitute the next major Big Data development, ultimately resulting in new and more complex business models.

Financial service provision

In the short term, financial service provision is primarily interested in efficiency: costs set against results. In this context, too, companies are fully aware of the fact that new providers, from other sectors for example, could develop new products on the basis of Big Data – products that are more finely tuned to the individual customer. Everyone is aware of the innovation potential that lies dormant in Big Data, but it is unclear who is going to be the first to actually make use of this.

Telecommunications and media

The telecommunications and media sectors have at their disposal a huge volume of data that enables them to provide better service to consumers. However, the challenge is to capitalize fully on this kind of information. Also, this framework demands more efficient internal processes. In the longer term, the automation of data analysis will lead to new and more complex business models.

Insurers

Experts in the insurance sector recognize the advent of current and future Big Data challenges. At present, handling personal data is governed by law on the one hand, and through consent from the individual on the other. In the future, other data sources may arise (as a result of monitoring and predictability, for instance), that could lead to the development of behavior-related insurance products. In this context we can think of data from telematics systems, social media, and electronic health records. Whether or not consumers support such innovative uses of data, and whether or not rules and regulations will be adjusted to respond, remains an open question.

Trade and retail

In trade and retail, Big Data analysis is clearly regarded as an opportunity to modernize. The infrastructure for data-intensive services will increasingly be offered as a resource. As a consequence, smaller organizations will also be able to benefit from Big Data with respect to segmenting and targeting. In the longer term, the data will become a product.

Market research

Market researchers, whether as a separate sector or as a subsector in larger organizations such as the credit-card industry or other financial or telecom-oriented service providers, are ideally suited to capitalize on the opportunities offered by Big Data. Algorithms, statistics and advanced visualization can be applied in order to present up-to-date and detailed market analyses and opportunities at any given moment. The development of this kind of data science is already rapidly rising, although data scientists are still very scarce. Fraunhofer IAIS is currently actively engaged in this project and offers a data science curriculum for data analysts, business analysts, marketers and financial personnel.

The Emerging Big Returns on Big Data 2013

In December 2012 and January 2013, TCS held a worldwide survey among 1,217 major organizations and published the findings in a report entitled *The Emerging Big Returns on Big Data*. The current overall picture of the field of Big Data can be summarized in the following five points. They represent the first stage of Big Data analysis in 634 organizations worldwide, over 12 economic sectors of which 83 percent have a turnover of more than 1 billion dollars.

1. More than half of the major organizations surveyed indicated that they developed Big Data initiatives in 2012.
2. Of these, 80 percent had improved business decisions as a result.
3. The majority of the data used remains structured, and comes from internal sources.
4. The greatest challenge within major organizations is ensuring that existing silos share their data.

5. In 2012, the median investment made by large organizations (not the average value, due to several substantial exceptions) was 10 million dollars.

We recognize that new Big Data-based insights can emerge in many areas of an organization: sales, logistics, marketing, HR and customer service. Expectations are highest for what new insights Big Data can bring to key concerns:

1. Identification of consumers who can generate most value.
2. Ability to monitor the quality of one's own products in detail.
3. Discovery of new needs among consumers, needs that current products cannot meet.

Question 3: How will these insights help me?

Answer: There are already concrete ROI expectations for eight corporate functions, and the focus and value of each function can be specified.

The TCS study predicted what the new insights will deliver. The results are set out below. We can easily recognize the twelve Big Data areas of the Fraunhofer IAIS, which we listed earlier in this report.

Corporate function	Highest value	Second-highest value	Estimated ROI over 2012
Logistics	Monitoring product shipment	Identifying peaks in costs	78%
Finance	Risk measurement	Forecasting	69%
Customer Service	Identifying consumer loyalty	Website behavior	56%
Sales	Identifying the most important consumers	Spotting cross-selling opportunities	54%
R&D	Monitoring product quality	Spotting new consumer needs	48%
HR	Estimating employee retention	Effectiveness of recruitment	48%
Operations	Product tracking	Planning deliveries	42%
Marketing	Measuring campaign effectiveness	Measuring media effectiveness	41%

Where can we expect a positive ROI with regard to Big Data?

Please note that an ROI approach is not always the best start to a Big Data project, especially if you wish to promote Data Discovery.

Question 4: What skills do I need?

Answer: BI professionals, marketers, sales people and IT management must collaborate with data scientists to fully master Big Data. Therefore recruitment and training are both essential.

In his 2012 book *Taming The Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics*, Bill Franks explains how organizations should set up and implement Big Data initiatives incrementally. Franks is the Chief Analytics Officer at Teradata, and also director of the Business Analytic Innovation Center of Teradata and SAS. Harvesting Big Data, as described in the figure on page 14, moves from internal to external and unstructured data in small increments. According to Franks, and many other Big Data experts agree with him, the quantity and the quality of insights gained will thus increase exponentially. He rightly emphasizes that an overall strategy is extremely practical and necessary to access the latent potential of Big Data. It is not only a technological issue – how do you blend structured and unstructured data? – but certainly an organizational issue as well: how do you transform the organization? Franks comments:

"The biggest value in Big Data can be driven by combining Big Data with other corporate data. By putting what is found in Big Data in a larger context, the quantity and quality of insights will increase exponentially. This is why Big Data needs to be folded into an overall data strategy as opposed to having a stand-alone Big Data strategy."

Franks emphasizes the paradox that, although organizations in traditionally data-analytical economic sectors can choose from a greater pool of suitable new employees, extra effort is nevertheless needed to keep up with the competition, due to that sectorial lead. Organizations in sectors without a data-intensive culture can gain a lead over their rivals though new analytical initiatives, but they may also have to reinvent the wheel themselves in some situations. In addition, several studies over the past few years have shown that the demand for Big Data analysts and data scientists easily exceeds the supply. As an illustration, here is a profile of a data scientist, so you can see just what kind of creature we're talking about.

Vacancy for Data Scientist on Monsterboard, April 2013

ING bank is seeking an innovative, creative and enthusiastic data scientist. Someone who has no difficulty with tackling and solving complex marketing issues with the aid of substantial amounts of data (Big Data) and modern IT resources. Someone who can stimulate and inspire colleagues to think "out of the box" and is himself/herself a classic example of this approach. [...] This data-driven marketing makes use of a voluminous data and application infra-

structure that is currently in the process of being completely renovated. In this framework, the focus lies on the establishment of a real-time, predictive infrastructure and a different way of (co-)operation. Aspects that are now at the forefront: Big Data, Netezza, Hadoop, Agile.

Function requirements:
The chosen candidate will be an academic with an informatics/mathematics/econometrics background, preferably with the accent on machine learning or artificial intelligence. The candidate may be currently busy with a PhD in one of these disciplines. Candidates should be familiar with (and, if possible, have experience with) technology and methodology in the field of structured and unstructured data (Netezza, Hadoop, MapReduce, Hive, R, SAS). They possess demonstrable knowledge and experience with statistical and mathematical methods of data analysis (such as neural networks, random forest, text-mining for example) and their application in business processes. In addition, candidates have roughly the following profile:

- *Are extremely inquisitive and continually searching for explanations*
- *Can cope comfortably with setbacks, are true stayers.*
- *Present sharp analyses – concise, to the point, convincing.*
- *Maintain a clear overview in complex, obscure situations.*
- *Are good in communicating, presenting and selling ideas.*
- *Are enthusiastic and genuine team players.*
- *Are quick to absorb new techniques and software.*

It is not our intention to place all the work on the shoulders of Big Data scientists. BI professionals, marketers, sales people and IT management must work with data scientists on new competences in order to fully master the complexities of Big Data.

The themes of sharing information and new technology are prevalent. In this setting, features such as expertise, including the ability to apply insight and prioritize, and willingness to contribute are major determinants. Attention must also be paid to developing trust between data scientists and business managers. Good data visualization is also important. Marketing and data experts want a clear business focus on the investments in Big Data, as well as on the various sorts of data that are needed for diverse issues. This is directly related to the previously mentioned availability of data from the silos in the business units.

Question 5: How do Big Data pioneers organize data management and IT processes?

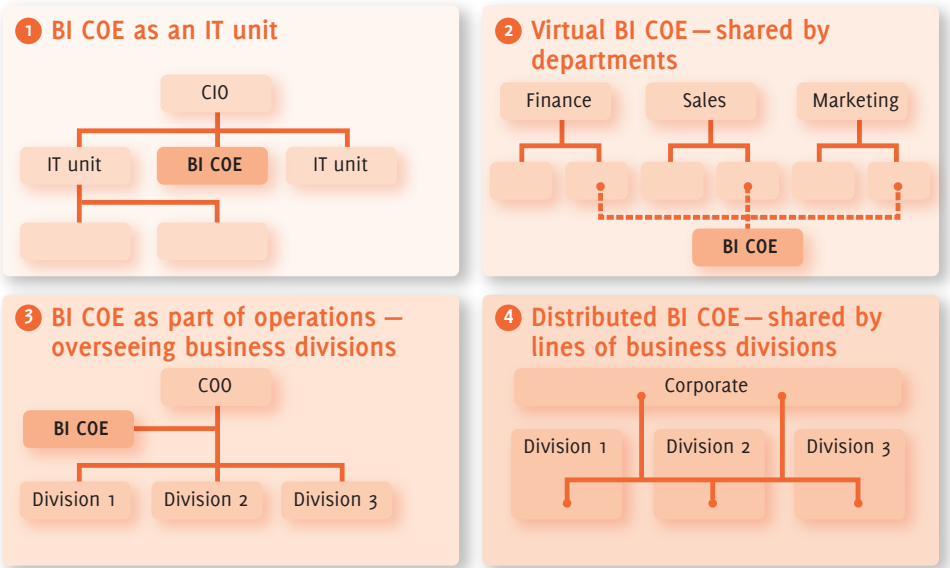
Answer: There are various ways to support the Big Data project; for example, by means of a so-called "next-generation" BI CoE (Center of Excellence). The most important way to provide support is by building successful collaboration between CMO, CIO and CDO (Chief Data Officer, if present), as is the case with Starbucks.

There is much commotion around the way in which organizations should shape their "real-time fact-based decision-making." Getting good results requires more than merely implementing technology such as Hadoop, Oracle Exadata, SAP HANA, SAS/ Teradata or IBM BigInsights. To start with, there must be change at the top. Anyone aspiring to capitalize on the potential of Big Data will have to pull some strings at the executive level. Depending on the source, the new executive function will be filled by a Chief Data Officer, Chief Decision Officer or the all-embracing Chief Digital Officer. What a Chief Data Officer and his/her Data Management Office actually do is reported in *The Role of the Chief Data Officer in Financial Services* by Capgemini. Topping the list of eighteen reasons to appoint a CDO, numbers 1, 2 and 3 mention the silo problem.

To generate good decision-making, business analysts must have various sorts of data at their disposal: data from the data warehouse, but also data from the web, Big Data and information from production systems, as well as input from partners and suppliers. Analysts spend more than half of their time acquiring this data. Because we generally wish to examine the facts as closely as possible, less time remains for solid analysis. For this reason, it is advisable to deliberate seriously and to continually evaluate the design and delivery of effective next-generation BI. Tools, datasets, knowledge and skills, and the right mindset in the organization are essential components of such structures.

It is common to concentrate the new core expertise in a (next-generation) Business Intelligence Center of Excellence (BI CoE), also referred to as a Business Intelligence Competence Center (BICC), and to coordinate it from this source. There are four main variants that are applied with businessmen, IT people and data experts, either separately or in a certain combination for a fixed time:

1. an IT unit that reports to the CIO
2. a virtual CoE, present in different functional business units
3. an operating unit that reports to the COO
4. a distributed CoE, present in the divisions and the overarching corporate organization.



Four ways to embed a next-gen BI CoE in the organization
Source: Kalakota (2012)

Please note: It is self-evident that the virtual CoE and the distributed variant will report to the CDO. Of course there must be good connections with the CIO office, which may be formally established without there necessarily being two captains on one ship or any inevitable clashes. The CIO office will be more engaged with the backend of the information systems, while the CDO and the CMO and their staff will explicitly direct their efforts to the acquisition of value from data for the consumer and the organization.

A remarkable work in this context is the study entitled *Big Data's Biggest Role: Aligning the CMO & CIO – Greater Partnership Drives Enterprise-Wide Customer Centricity* by SAS and the CMO Council, which was published in March 2013. Due to the increasing focus on Big Data analysis, the CMO and the CIO Office are now growing toward one another, whereas a wide gap was clearly evident not so very long ago.

Other handy tips for steering the proliferation of data and the necessary data integration in the right direction (in short: how we design the next-gen BI) are presented in the report *Big Data Analytics: Future Architectures, Skills and Roadmaps for the CIO*, by IDC and SAS. (In the figure below we have streamlined the phrasing and the functional elements in the report.)

	Maturity			
	Stage 1	Stage 2	Stage 3	Stage 4
Data Governance	Little or none (Skunk words)	Initial data warehouse model and architecture	Data definitions and models standardized	Clear master data management strategy
Technology & Tools	Simple historical BI reporting and dashboards	Data warehouse implemented, broad usage of BI tools, limited analytical data marts	In database mining, and limited usage of parallel processing and analytical appliance	Widespread adoption of appliance for multiple workloads Architecture and governance for emerging technologies
Staff Skills (IT)	Little or no expertise in analytics – basic knowledge of BI tools	Data warehouse team focused on performance, availability, and security	Advanced data modelers and stewards key part of the IT department	Business Analytics Competency Centre (BACC) that includes "data scientists"
Staff Skills (Business/IT)	Functional knowledge of BI tools	Few business analysts – limited usage of advanced analytics	Savvy analytical modelers and statisticians utilized	Complex problem solving integrated into Business Analytics Competency Centre (BACC)
Financial Impact	No substantial financial impact No ROI models in place	Certain revenue generating KPIs in place with ROI clearly understood	Significant revenue impact (measured and monitored on a regular basis)	Business strategy and competitive differentiation based on analytics

Loosely taken from IDC & SAS (2011)

The figure presents, moving from top to bottom, the coherence of adequate data governance, the right technology and tooling, and staffing in the IT and business/IT department. The consequence – the anticipated impact on the financial results of the organization as an accepted indicator of the overall business performance – is displayed at the bottom. The increase in maturity is shown in four stages from left to right, moving from minimum to optimum. It is up to you to use the information provided in this report to determine where your challenges and opportunities lie, and what is the most suitable route for you to take toward growth.

CIO and CDO a successful team at Starbucks

Organizations do differ. Big Data may provide an impulse for the CIO and the CMO to collaborate more intensively, but also a newly appointed CDO (Chief Data Officer) and the CIO can form an excellent team, as is the case at Starbucks.

Starbucks is worth 13.3 billion dollars, and is extremely successful with its combination of brand stores and digital ventures such as social media and mobile networks. A CDO was appointed in March 2012 and he now supervises 110 employees. He works in close cooperation with the CIO, who was appointed at the same time and whose office comprises 760 employees. The great change in the first year was the repositioning of technology and digital services by orienting them more toward the customers and the shops.

All Starbucks' digital projects have now been combined: web, mobile, social media, digital marketing, loyalty programs, e-commerce, wifi, the Starbucks network and new shopping technology. In bygone days, worldwide digital marketing, the Starbucks cards and mobile payments, and loyalty programs were three separate units. These have now been successfully integrated. The CIO and the CDO work in harmony and the entire digital Starbucks roadmap is thoroughly examined on a weekly basis. The traditional, separate Centers of Excellence have been replaced by intensive project teams, working in rapid iterations (*MIT Sloan Management Review*, 2013).

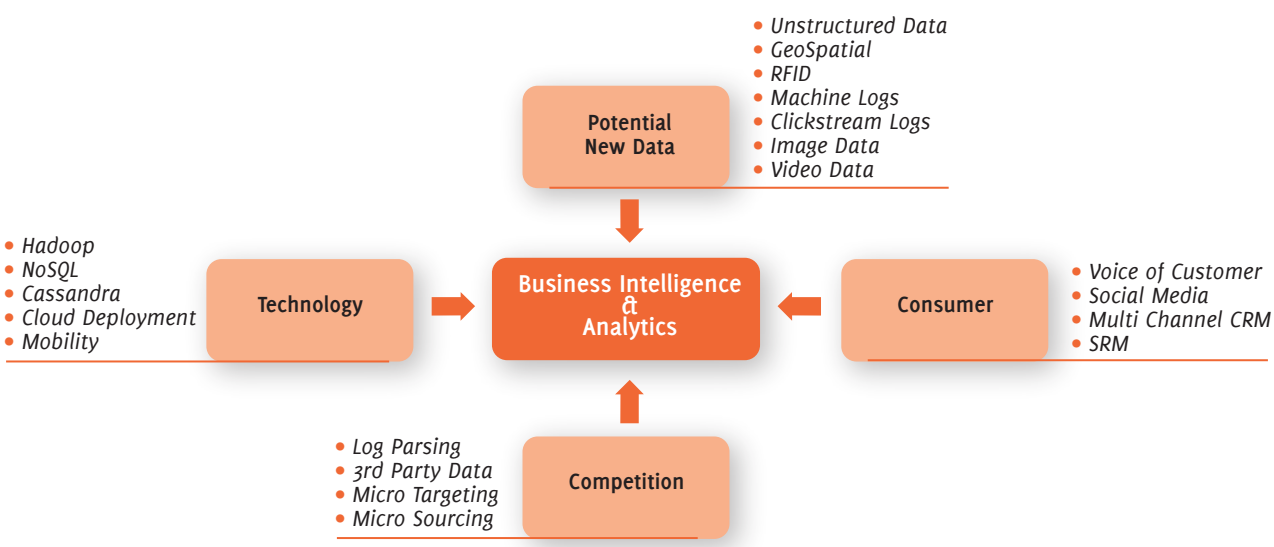
Question 6:
How can I merge my structured and unstructured data?

Answer: Business Intelligence must become faster and more analytical. Ideally you would use a modern, integrated information architecture with specialist hardware, a high-speed network, and in-memory analytics for this purpose.

For ten years, stretching from 2002 to 2012, Sam Palmisano was CEO of IBM. In January 2010, he made his "Decade of Smart" speech. In the course of the second decade of the 21st century, Palmisano predicted that we will find an increasing number of answers in the rapidly growing digital dataflow that we have at our disposal.

Palmisano deliberately avoided talking about Big Data. All conceivable data taken together – large, small, structured or unstructured, relational or graph, metadata and masterdata, in ontologies and taxonomies – will jointly lead to new insights and better and quicker decisions.

Here too, the *devil lurks in the details*. In our humble assessment, everything that is currently referred to as "Business Intelligence" must dig a bit deeper or shift up a couple of gears. In other words, everything must become more analytical and faster. Krish Krishnan, who published *Building the Unstructured Data Warehouse* with data guru Bill Inmon at the beginning of 2011, and whose new book *Data Warehousing in the Age of Big Data* recently appeared, presents the new challenges in the field of BI schematically, as in the following figure.



Source: Krishnan (2012)

A more thorough analysis in the context of data integration will be presented shortly, but the impact will be evident. In his overview entitled "Big Data & Analytics," from which the above figure showing the data challenges to BI has been borrowed, Krishnan sums up the central developments, challenges, solutions and limitations that affect Business Intelligence in the framework of Big Data: the inclusion of all kinds of unstructured and external datasets in the analyses and insights for the organization. In its generality, Big Data is akin to the more traditional enterprise data domains in the organization, as is outlined in the following Oracle table:

Data Realm	Structure	Volume	Description	Examples
Master Data	Structured	Low	Enterprise-level data entities that are of strategic value to an organization. Typically non-volatile and non-transactional in nature.	Customer, product, supplier, and location/site
Transaction Data	Structured & Semi-structured	Medium-High	Business transactions that are captured during business operations and processes	Purchase records, inquiries, and payments
Reference Data	Structured & Semi-structured	Low-Medium	Internally managed or externally sourced facts to support an organization's ability to effectively process transactions, manage master data, and provide decision support capabilities.	Geo data & market data
Metadata	Structured	Low	Defined as "data about the data". Used as an abstraction layer for standardized descriptions and operations (e.g. integration, intelligence, services).	Data name, data dimensions/units, definition of a data entity, or a calculation formula of metrics
Analytical Data	Structured	Medium-High	Derivations of the business operation and transaction data used to satisfy reporting and analytical needs.	Data that reside in data warehouses, data marts, and other decision support applications
Documents and Content	Unstructured	Medium-High	Documents, digital images, geo-spatial data, and multimedia files.	Claim forms, medical images, maps, video files
Big Data	Structured, Semi-structured & Unstructured	High	Large datasets that are challenging to store, search, share, visualize, and analyze	User and machine-generated content through social media, web and software logs, cameras, information-sensing mobile devices, aerial sensory technologies, and genomics

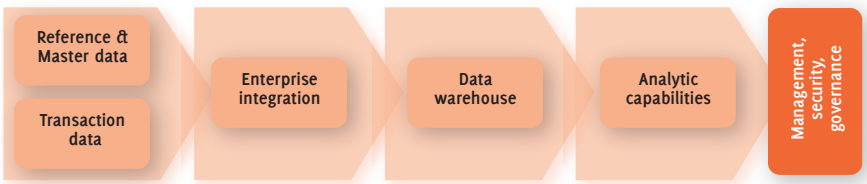
Source: Oracle (2012)

Altogether, this requires ongoing development of the information architecture, which we should envisage – according to the Oracle – in three steps:

Step 1

This is the traditional situation with only structured data:

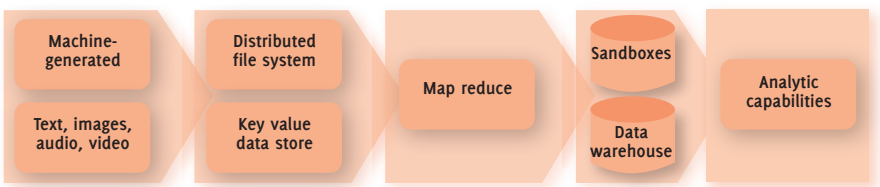
Structured



Step 2

We add unstructured (Big) data to this model:

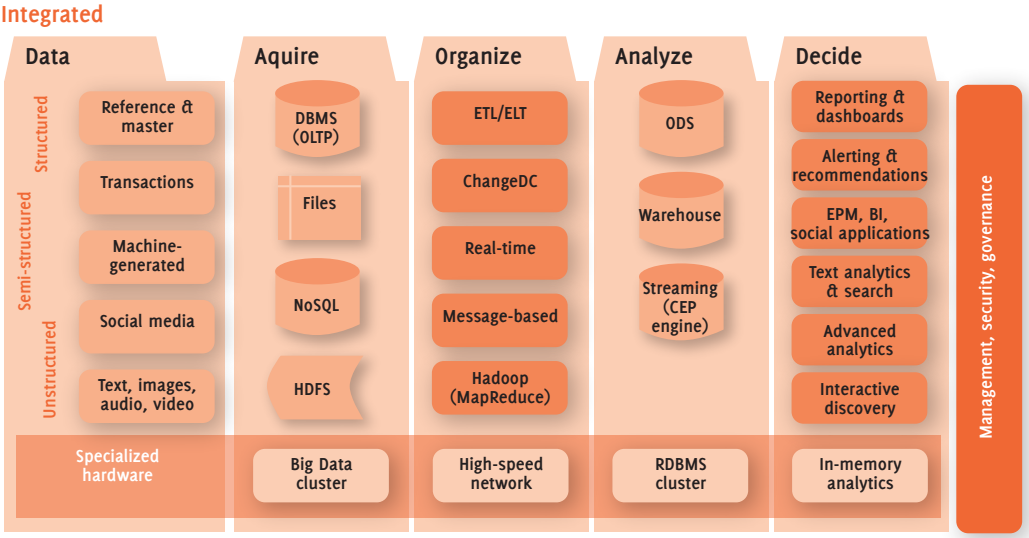
Unstructured



Step 3

The combination of 1 and 2 must be able to be appropriately scaled. Accordingly, a modern, integrated information architecture has a setup similar to the following, complete with specialized hardware, a high-speed network and in-memory analytics, among other features:

Integrated



Question 7: Which new technologies should I be watching?

Answer: the following three aspects of the technology behind Big Data are crucial points of concern for those involved in business IT: a) the three routes toward data integration (Hadoop, NoSQL and hybrid), b) ten Big Data core technologies for the entire process, ranging from data capturing to visualization, and c) the scalability of Big Data on the basis of the three Vs.

Routes to data integration

Some data integration solutions only work with the Hadoop platform, whereas others can cope with all kinds of Big Data stores and also offer analytics to boot. Organizations must be perfectly clear which data stores are important to them, and should align their data integration products and processes to these. Richard Daley, the co-founder of Pentaho, supplements this as follows, giving a good initial picture of what Big Data involves in technological terms. Daley distinguishes between Hadoop integration, the NoSQL route, and hybrid Big Data integration.

1 Hadoop integration

Many Big Data initiatives are based on Hadoop variants such as Apache Hadoop, Cloudera, MapR, Hortonworks and Amazon Elastic MapReduce. NoSQL databases are used to a lesser extent, as are high-performance relational analytical databases such as Greenplum, Infobright, Netezza, Teradata, Vectorwise and Vertica, among others. In a formal technical sense, Hadoop is actually a form of NoSQL.

Even the most basic Hadoop integration solutions go a step further than Hive integration these days. They support the entire range of Hadoop interfaces such as MapReduce and HDFS, and Hadoop eco-systems such as Pig, Sqoop and Oozie.

- Apache Hive is a powerful data warehouse application for Hadoop by means of which data can be accessed via Hive QL, which is a language similar to SQL.
- Apache MapReduce is the software framework for Hadoop, enabling the parallel processing of large quantities of data on large computer clusters.
- Apache Hadoop Distributed File System (HDFS) is the storage system for Hadoop applications. HDFS reproduces data and distributes the results to the computing nodes of a cluster for reliable and extremely high-speed processing. Some Hadoop distributions offer alternatives to HDFS, such as NFS (from MapR) and Apache Cassandra (from DataStax). These give even better performance and can cope with data under Windows, OS X and Linux.
- Apache Pig is a script language for MapReduce programs under Hadoop. Due to their structure, Pig programs are easy to run in parallel, so that they can deal with extremely large datasets without difficulty.

- Apache Sqoop is a tool enabling the easy migration of bulk data between Apache Hadoop and structured datastores, such as relational databases. A simple command-line tool, Sqoop (SQL-to-Hadoop), positions tables or even complete databases in HDFS files, generates Java classes for processing, and makes it possible to place information from SQL databases directly into the Hive warehouse.
- Oozie is a server engine allowing Hadoop to run workflow jobs, such as MapReduce, Pig, Hive, Sqoop, HDFS operations and subworkflows.

2 The NoSQL route

If you opt for the NoSQL database route instead of Hadoop (although in technical terms Hadoop is also a form of NoSQL), your Big Data solution will have to support MongoDB, Cassandra and HBase. MongoDB is rapidly gaining in popularity and offers an easily accessible, scalable, high-performance, open-source NoSQL document store.

A NoSQL data-integration solution must provide the following facilities:

- *A user-friendly visual development environment* that every IT worker, data analyst and business user can deploy to manipulate, visualize and explore data and to report on his/her activities.
- *Hybrid data integration*, so that the NoSQL database can be directly used as a source for reports and dashboards. Other data must also be able to be added to a data warehouse to allow a 360-degree view of the business.
- *Integration of the NoSQL database* with existing Big Data and traditional data stores to give a complete data management and data analytics solution.

3 Big Data integration with Hadoop and NoSQL

Instead of focusing on data integration through Hadoop or NoSQL, you can combine both in order to link up to various Big Data sources and enterprise stores. This kind of overarching integration must perform four tasks:

a. alleviate the limitations of Big Data storage and processing, so that Hadoop is no longer subjected to delays when attempting to gain access to data in computer clusters, enabling unimpeded implementation of sort, group and join queries on NoSQL databases.
b. remove technical barriers. Users need simple, user-friendly visual development interfaces for high-performance data input, output and manipulation, regardless of whether they are working with the Hadoop or NoSQL Big Data platform. IT workers, developers, data scientists and business analysts must be able to tap into, integrate and analyze both Big Data and traditional data without undue difficulty.
c. guarantee integration with enterprise data. The Big Data platform must be integrated with enterprise datastores. Therefore there must be good tools to connect Hadoop and NoSQL databases to traditional relational databases, exchange formats, and enterprise applications.
d. provide a complete business-analytics solution with reports, dashboards, interactive visualization and exploration, and predictive analytics.

Ten Big Data core technologies

The processing stages that apply to Business Intelligence applications also apply to Big Data, but demand extra technological effort to enable the complete process of data capturing, storage, search, sharing, analytics and visualization to occur smoothly. In his book entitled *Big Data Glossary*, Pete Warden discusses a total of sixty technological innovations and provides the following concise overview of Big Data concepts and tools.

1 Data acquisition

(such as Google Refine, Needlebase, ScraperWiki, BloomReach, for example)
For accessing various data sources, internal or external, structured or unstructured. Most interesting public data sources are poorly structured, full of contamination and difficult to open.

2 Serialization

(such as JSON, BSON, Thrift, Avro, Google Protocol Buffers, for example)
At various points during processing, the data will be stored in files. These operations all require some sort of ranking.

3 Storage

(such as Amazon S3, Hadoop Distributed File System, for example)
Traditional file systems are unsuited to large-scale, distributed data processing, but nevertheless the Hadoop Distributed File System is actually suited to this function.

4 Cloud

(such as Amazon EC2, Windows Azure, Google App Engine, Amazon Elastic Beanstalk, Heroku, for example)
Hiring computers as virtual machines in a cloud environment is increasingly becoming standard procedure. In this way, you can make use of large processing capacity for Big Data applications at relatively low cost.

5 NoSQL

(such as Apache Hadoop, Apache Cassandra, MongoDB, Apache CouchDB, Redis, BigTable, HBase, Hypertable, Voldemort, for example. See http://nosql-database.org for a complete list)
NoSQL (Not only SQL) is a broad class of management systems that deviates from the classical relational model.

6 MapReduce

(such as Hadoop and Hive, Pig, Cascading, Cascalog, mrjob, Caffeine, S4, MapR, Acunu, Flume, Kafka, Azkaban, Oozie, Greenplum, for example)
In traditional relational database environments, all processing takes place via a special query language after the structured information has been loaded. In contrast, Map-

Reduce reads and writes unstructured data to all sorts of file formats. The interim results are passed on as files, and the processing is divided among many machines.

7 Processing
(such as R, Yahoo! Pipes, Mechanical Turk, Solr/Lucene, ElasticSearch, Datameer, Bigsheets, Tinkerpop; start-ups: Continuuity, Wibidata, Platfora, for example)
Filtering concise, valuable information from an ocean of data is a challenge, but there are already many solutions that can help with such tasks.

8 Natural Language Processing
(such as Natural Language Toolkit, Apache OpenNLP, Boilerpipe, OpenCalais, for example)
Natural Language Processing extracts meaningful information from untidy, human-based language.

9 Machine Learning
(such as WEKA, Mahout, scikits.learn, Skytree, for example)
Machine Learning systems automate and optimize decision-making. The recommendations given by Amazon, for instance, are well-know applications of this function.

10 Visualization
(such as GraphViz, Protovis, Google Fusion Tables, Tableau Software, for example)

This is one of the best ways to extract significance from data. Thanks to interactive graphics, the presentation and exploration of information blend together.

Scalability of Big Data

In conclusion, let us examine how we can classify Big Data on the basis of the well-known basic trio of *Volume*, *Variety* and *Velocity*. To establish a picture of Volume and Velocity, it is interesting to check how long data transport actually takes from one location to another through a landline. Many people, used to normal domestic usage, have simply no idea of the time involved. Through a T-1 line, it takes 90 minutes to send 1 gigabyte from A to B. With Thin Ethernet it takes 14 minutes, and with Fast Ethernet 1 minute. This is not hyper-fast, but it seems reasonably acceptable. In contrast, 1 terabyte through a T-1 line takes almost 66 days. With Thin Ethernet this is more than 10 days, and with a Fast Ethernet connection it still takes 24.5 hours. Even with Gigabit Ethernet this still costs almost 2.5 hours. For high-speed Data Discovery iterations such as the 80,000 instances of the Capital One credit card company, this kind of bottleneck is unacceptable, and this applies to other scientific environments such as those of the Netherlands Forensic Institute or a genome institute. A thousand genomes amount to 200 terabytes, and we are still only speaking about the volume aspect of the matter.

In their overview article entitled "Tackling Big Data," Michael Cooper and Peter Mell of the IT Laboratory Big Data Working Group of NIST, the American National Institute for Standards and Technology, distinguish three Big Data types. They are related to the three Vs in the following way:

Volume	Velocity	Variety (semi-structured/ unstructured	Requires Horizontal Scalability	Relational Limitation	Big Data
No	No	No	No	No	No
No	No	Yes	No	Yes	Yes, Type 1
No	Yes	No	Yes	Maybe	Yes, Type 2
No	Yes	Yes	Yes	Yes	Yes, Type 3
Yes	No	No	Yes	Maybe	Yes, Type 2
Yes	No	Yes	Yes	Yes	Yes, Type 3
Yes	Yes	No	Yes	Maybe	Yes, Type 2
Yes	Yes	Yes	Yes	Yes	Yes, Type 3

Type 1 *Non-relational data representation required for effective analysis*
Type 2 *Horizontal scalability required for efficient processing*
Type 3 *Non-relational data representation processed with a horizontally scalable solution required for both effective analysis and efficient processing*

In other words: ***the data representation is not conducive to a relational algebraic analysis.***

Source: Cooper & Mell (2012)

Question 8:
What is looming on the horizon?

Answer: With regard to your transformation and performance, you should aspire to belong to the category of the "Digirati," in line with the study published by MIT and Capgemini.

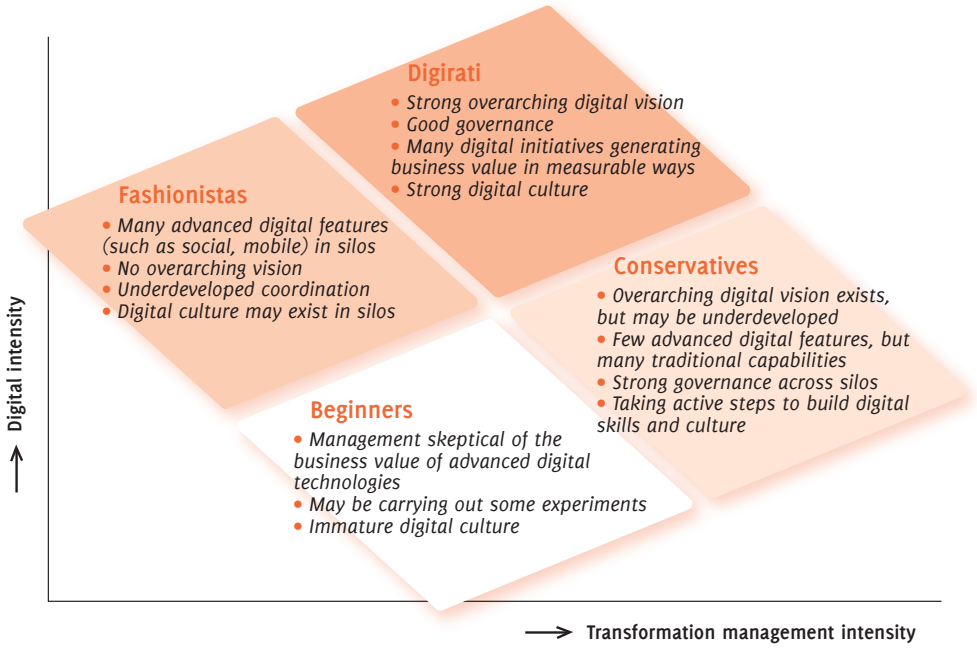
Big Data is a part of the digital path you are following. And there is good news about that path. As we mentioned in the Introduction, there is evidence that a digital strategy is rewarding in terms of turnover, margin and shareholder value. Such indicators are less relevant to the public sector, but here, too, it is worthwhile investing wholeheartedly in digital options. In May 2013, the Dutch government for instance announced that all government services must be fully digitized by 2017. This should lead to annual savings of 300 to 400 million euros.

The potential returns of your digital activities, and thus now of Big Data in particular, is emerging in relevant studies. The results of recent research by the Massachusetts Institute of Technology (MIT) and Capgemini, as well as 391 other organizations, establishes a starting point for a longer-term strategy, a vision looming on the horizon of how you could evolve from a viewer into a top performer.

This study is called *The Digital Advantage: How Digital Leaders Outperform Their Peers in Every Industry* and it distinguishes four archetypes of digital intensity on the one hand and transformative capabilities on the other. The true "laggards" are politely referred to as "Beginners." They do not really believe in the power of digital technology, and perform a few half-hearted experiments at most. Their real problem is that their ambition to change is insufficiently developed.

This appraisal does not apply to the "Conservatives," the group that is depicted next to the "Beginners." Their transformative capabilities are actually high, and this is expressed in the way in which this group regards digital technology. The "Conservatives" do have a vision for digital technology, but they really ought to be more enthusiastic. In contrast, their business sections are well managed, and work is genuinely being done with the aim of realizing a better digital culture. However, there is almost no advanced IT in the departments.

This is the reason why we have depicted the *Digital Advantage* quadrant as a diamond here. The "Conservatives" are better off than the "Beginners," as the *Digital Advantage* study clearly shows.



The performance of organizations that exhibit high digital intensity and good transformative capabilities overshadows that of their peers. Top performers are the so-called "Digirati." They have a strong view of digital affairs and a digital culture. This type of organization is well managed and there are (therefore) many digital initiatives that palpably add value.

The remaining category, the so-called "Fashionistas," is carried along on the waves of fashion, as the name indicates. From the outside, this type of organization has a reasonably attractive digital profile, in the social media and in mobile applications, but the initiatives are fragmentary, a true overarching digital vision is lacking, and collaboration with the organization leaves much to be desired. This being the case, the "Fashionistas" may indeed be hip, but it could all be a good bit more efficient and sincere, certainly with regard to their transformative skills.

The interesting aspect of these kinds of archetype is that you will never actually find an actual organization that is identical in all aspects with one described in the model. An archetype is a reference category. The significance of archetypes is that we are compelled to look closely and candidly in the mirror and investigate just who we are in certain respects, and how we can perhaps change for the better. This makes the *Digital Advantage* study a useful aid when formulating a plan to determine a strategy aimed at obtaining more benefit from Big Data. The study does not place the emphasis on Big Data, as already mentioned, but rather on solid data integration and the importance of a data strategy.

Looking to the horizon and transitional stages

Embracing "Digirati" characteristics can help you get your Big Data initiatives off the ground. That involves embedding Big Data in a greater body of digital strategy and strategic necessity. The study provides the following four transitional stages as intermediate targets:

1 Working on a transformative vision

Imagining how the organization might look in the future is a way to change the organizational mindset. On the one hand, it makes clear which rules will no longer be valid in the future and, on the other, it offers a glimpse of a new way of working. Enough books and aids (such as TheIdealFrame.com, for example) have been written about framing ideas, with the goal of providing a positive outcome. Greg Sattel gives an example of how to frame a business transformation:

> *"[...] business models can no longer be treated as stone tablets divined by wise men on mountains to last for eternity."*

And another one, formulated by Clay Shirky, is quoted by Kevin Kelly:

> *"Institutions will try to preserve the problem to which they are the solution."*

2 Digital governance is an important trump card

To make it more certain that investment in Big Data will be successful, it is necessary to examine the organizational governance structure, particularly with reference to next-generation Business Intelligence. The text at question 5 provides some reference points for this examination.

3 Members of staff must be involved and must remain so

Only by involving employees at an early stage can a Big Data integration strategy be truly successful. An internal innovation jam for ideation, implementation, and execution of Big Data innovations has proven to be a suitable instrument for this.

4 Collaboration between business and IT

Ensure that teams from business and IT work together to actively execute data integration.

Intensive data integration requires the adaptation of architectures and processes. For example, marketing can direct its campaigns to smaller segments thanks to localized data. In hospitals, analyses of DNA deviations that initially took two years to perform can now be done in a few weeks. Methods of diagnosis and treatment are thus altered.

Question 9: What does this mean in organizational terms?

Answer: Actually making your organization data-driven begins with a good strategic plan to enable data, analytics, tools and people to jointly create business value. The directors, technology professionals, data scientists and managers must jointly determine where the greatest profits are to be obtained, and they must then select the first projects.

Let us take a concrete example. In his 2006 bestseller *Marketing Genius*, business guru Peter Fisk adeptly explained how most organizations deal with data. Fisk thought it foolish that everyone is constantly seeking more data. Most of the information that organizations accumulate, according to Fisk in 2006, is completely worthless. The real issues remain unaddressed as the data is aggregated and averaged much too quickly, so that all relevant information evaporates. It is often the case, Fisk argues, that management has already decided on its plans and it makes no difference whatsoever what the analyses show:

> *"Most organizations are weighed down by research reports, tracking data, analytical spreadsheets and the like. The research industry is consequently huge too, perpetuating the myth that more data is good for you. Yet most information collected by organizations is useless. It doesn't address the issues important to them, it is quickly aggregated and averaged so that any useful knowledge is smoothed out, and it is more often than not requested by managers who already decided what they want to do, irrelevant of the research findings."*

You certainly do not need to be a marketing genius to recognize the futility of the practice Fisk has identified. If this is to become Big Data practice, it should only be used to support decision-making in a formal sense, as it offers no solid advantage. This is grist to the mill of the cynics who claim that BI has not kept its promise and that Big Data is following in BI's slipstream. If new insights into structural innovation cannot get anywhere, continuing to collect data is no more than an empty ritual that should be halted. Fisk candidly commented that you should know what you are engaged in and which questions you want to have answered:

> *"The first steps in achieving insight are to stop [...] collecting more data than you need, resisting the desire to research everyone constantly, and to ask every possible question. There is also a temptation to jump into research without clarifying objectives, often finding that it has no particular purpose, or cannot answer the most important questions."*

In his caricature, which undoubtedly bears a grain of truth and was intended to provide a wake-up call to organizations, the so-called "plan" that management cobbled

together and intended to implement was not formulated in a methodical way. There is no real mention of a direction and route that deserves the epithet of "plan."

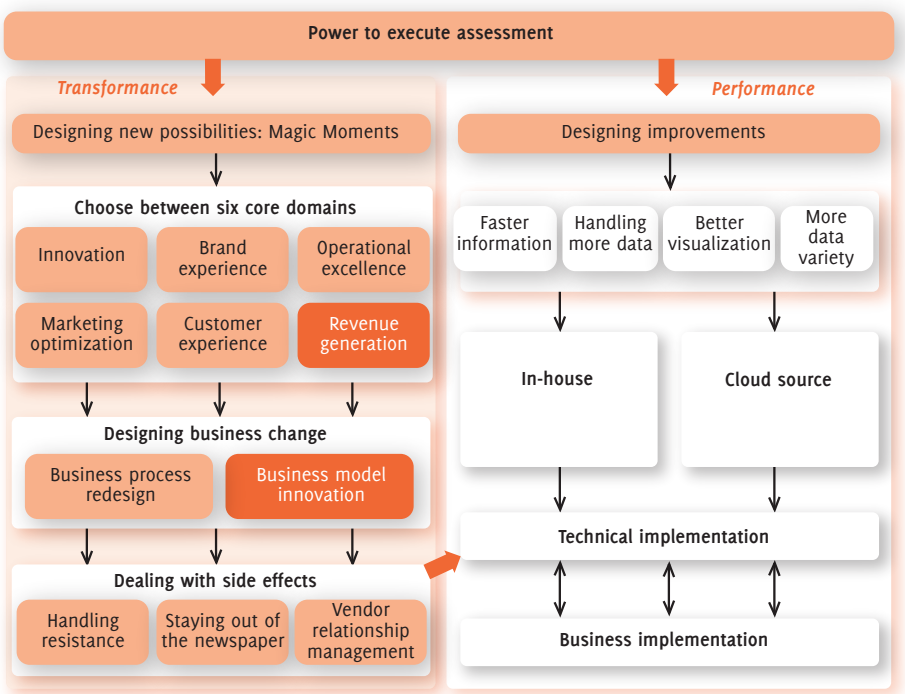
In March 2013, seven years after Fisk's warning and despite evidence that Big Data initiatives, if well applied, easily repay themselves, McKinsey expressed more or less the same sentiment. Most organizations had no clear plan to become truly data-driven, although those who do seriously work on this development soon achieve easily 5-to-6 percent more productivity and profit than the competition.

Making your organization data-driven begins with a good strategic plan of how data, analytics, tools and people can jointly create business value. This type of plan is, according to McKinsey, the common language of the directors, technology professionals, data scientists and managers when determining where the greatest profit can be achieved and when selecting the two or three data-intensive projects with which you wish to start.

With regard to Big Data, McKinsey draws a parallel with the situation forty years ago. Well-formulated strategic plans were more of an exception than the rule at that time. The pioneers booked striking results through these plans and an increasing number of companies slowly began to see the light.

Keeping this in mind, it is advisable to determine your execution capability, as in the drilldown that VINT proposes in the following figure, and allow yourself to be led by the lines of transformation and performance. Following these proposals, your organization will automatically come face to face with the most important issues that you will have to deal with.

Exploring Big Data potential



This drilldown gives an overview of the broad contours of the organizational aspects. We wish to elucidate four particular aspects here.

Magic moments and practical improvements

The true "Digirati," the companies that have already had much more experience with crunching unstructured data, are primarily seeking improvement in performance: transporting data through a line from A to B more quickly, better visualization, integrating even more external and internal data, etc. These companies will also be more willing to look to cloud solutions. Organizations that are largely engaged in exploration begin by creating magic moments: a creative start-up stage in which ROI thinking is taboo for the present and long-protected customs are abandoned in order to arrive at genuinely transformative ideas.

Risks

In the lower left-hand corner of the figure you will see that there are side effects that you will have to take into account: the risk that you may be portrayed negatively in the news, for example. Whatever the case, *Privacy by Design* (see our third Big Data report) offers you the necessary guidance. However, resistance within the organization – because you are going to manage processes differently in the future, for example – is also a potential side effect. Vendor relationship management, the reverse of customer relationship management, is also a potential danger. If common citizens and consumers can gain more control over their own data, business models will arise in which these groups will ultimately – and more often – be able to claim their just rewards.

Strategic choices

In our first research report, entitled *Creating Clarity with Big Data*, we presented six general areas where organizations could readily begin with data integration and Big Data. These are:

- innovation
- customer experience
- brand health
- marketing optimization
- turnover generation
- operational efficiency.

It is self-evident that we should begin at the spot that is most important to the organization: at the heart of the business and in compliance with the strategy. To Amazon, for example, this is customer experience. On top of the mountain of data, the algorithm that also recommends books to you is an expression of this strategy. To Shell this "innovation" lies in the tracing and development of energy sources. In the Netherlands, the National Forensic Institute has operational efficiency as its the impetus behind its chain of investigation. Equens, the largest pan-European payment processor, started up relatively recently and has chosen to use Big Data for turnover

generation, by means of which it differentiates its earnings model. An exception may consist of initiating Big Data projects for general use, such as some telcos and organizations such as TomTom have done. The idea behind this course of action is to cultivate understanding for the Big Data activities of these companies.

Innovation at Shell

For a company such as Shell, innovation is an extremely important area where much effort and investment are devoted to Big Data. To an increasing extent, it is becoming difficult for Shell to access conventional sources of oil and gas. Seismic research produces 50 to 100 petabytes of data. In bygone days, a maximum of a few thousand sensors were deployed, but nowadays this may be up to a million. In a new unit, called "Technical and Competitive Information Technology" (TACIT), work is being carried out on new techniques for data visualization in 3-D and 4-D. This unit gives substance to the strategic intention to become the most innovative energy company. Shell foresees that it will need a hundred times more computer power than it has at present in order to reach its goals. There is a great deal to play for. A difference of thirty meters in a deep-sea drilling venture can determine whether or not oil may be found. Each drilling costs 100 million. In that case, it is advisable to have the very best analytics at your disposal, as well as the machines that can process and visualize the data. In TACIT, Shell collaborates with MIT, Stanford, Ferrari, the Chinese Academy of Sciences, HP and Intel, among others.

Operational efficiency at the NFI

The Netherlands Forensic Institute (NFI) has to deal with petabytes of data annually. The request it often receives is whether or not it can distill criminal behavior from a large variety of data within a relatively short time. This may range from a Wordfeud conversation to data on a damaged harddisk to data in a TOR network. In short, there is an enormous mountain containing all kinds of snippets of information, from all over the place. Time also plays a major role in the formula. In the meantime, the NFI has become so quick in tracing child pornography, for example, that the judicial system is now beginning to suffer from congestion. The speed with which the NFI can make a statement is often essential, for sometimes a suspect has to be released after a maximum of 72 hours in custody.

The amount of data and data sources to be examined in criminal cases, mostly in fraud, murder and child pornography cases, is increasing exponentially. To enhance the effectiveness of this investigation, the NFI has developed an advanced software application, Xiraf. This application can analyze large quantities of data at a high speed and make them searchable. A user, such as a police detective, can open the tool via a secure internet link and enter a search term.

There is also the Bomb Data System (BDS), a digital database that contains all the data on incidents and threats with explosives in the Netherlands. This state-of-the-art system was commissioned by the National Coordinator for Security and Counterterrorism. The BDS enables the storage of information on explosives and incidents with chemical, biological, radiological and nuclear substances in the Netherlands, and these can be directly digitally shared with other users. The BDS is also accessible via smartphone or tablet and therefore can be used at the scene of a crime. The speed of information processing is crucial in the detection process, in the field in the case of a direct threat, for example, but also in national trend and threats analysis and for forensic objectives.

New Equens data conversion cancelled

Until recently, Equens, the largest pan-European processor of payment data, had never asked itself if it was possible to do more with the data at its disposal. After consultations with lawyers and discussions with the banks, Equens set up a service that intended to extract more value from its data. Encrypted and anonymized at card level, these data provide insight into spending at various identifiable locations. Real-time expenditure can be registered at shop level, in postal code zones and according to purchasing patterns. Simple questions that Equens could answer with a push of a button are, for example:

- What is the average expenditure of a consumer in a certain shop?
- In which other shops do my customers also spend their money?
- What were the expenditures in specified postal code zones yesterday?

These are the first steps toward innovating the earnings model. The approach is extremely cautious – *Privacy by Design* right down to the finest details – and ideas about new applications will not be hastily implemented but first thoroughly examined and discussed with experts and stakeholders. Nevertheless, the announcement of these plans did lead to some disquiet. The Dutch Minister of Finance and Eurogroup President Dijsselbloem even mingled in the debate and Equens announced that it would halt these activities for the present. Only if there is broad societal support will the theme be revived. Despite the fact that there seemed to be possibilities in a legal sense, the pressure to suspend the activities was overwhelming.

Question 10:
How does this affect everyday life?

Answer: In technological terms, questions 6 and 7 make the difference with traditional RDBMS environments, but a purposeful data focus on the combination of business, organization, and technology is both the core and the aim.

Conclusion and checklist

You are currently developing your Big Data potential of 2020 and, for this reason, this fourth and final VINT Big Data research report concludes with a checklist of twenty questions. They have been taken from our Big Data reports and from all interaction with the experienced experts with whom we have had contact online or in person. The issues of whether or not you are properly situated to tap into the Big Data potential, and of where the possible challenges may lie, can be answered using this checklist.

As a summary of this report and as an overture to the checklist, we first present the seven most important conclusions and recommendations in the mixture of business, organization and technology.

1 Structurally exploiting Big Data is becoming affordable

Only a few more years and the worldwide production of data will reach a stunning 40,000 exabytes, while the costs of working with that data will have diminished to 5 percent of current costs. Every sector and every organization can benefit from this.

2 Develop digital and organizational competences

Big Data potential lies in the meaningful combination of unstructured and structured data. Much work still has to be done on the digital side, but management must also do better. The Big Data potential of organizations lies at the interface of new digital and organizational competences.

3 Data Discovery is the next-generation BI

Insights from the new dataflows can cause a great deal of upset. This is one of the reasons why the concept of "digital transformation" is mentioned so frequently in relation to Big Data. Patterns become manifest and Data Discovery becomes a core component of the next generation of Business Intelligence.

4 Your aim is to belong to the Digirati category

Some organizations are better situated to realize their Big Data potential. They have been working toward a vision looming on the horizon for a much longer time, with the aim of making the organization fit for a digital future. We refer to them as "Digirati," and have given a description of their character traits in the section on question 8.

5 Ensure you have sufficient Big Data knowledge

Technological and analytical knowledge of processing unstructured data is a must in order to mix information from external and internal sources with structured data.

6 Big Data is "magic moments" and performance

Transformative Big Data initiatives begin with "magic moments": by choosing a domain in which your organization wishes to excel, while taking into account the risks and side effects. Performance Big Data initiatives are directed to existing projects with the aim of improving the performance. They involve the choice of what you yourself wish to do, and what you are going to (cl-)outsource (see question 9).

7 Big Data is an organization-wide evolution

In summary: you must develop a vision, acquire the right technological and data-scientific competences for the organization, give intensive data focus to the existing operation, and arrange your (data) governance properly.

In other words, you ought to read the following text attentively. We began with it, and it is appropriate to follow this report with it:

Big Data is the new, intensive, organization-wide focus on business, organization and technology. Accordingly, you must make sure your organization has sufficient technological and analytical knowledge, as well as the appropriate digital and organizational competences. Your goal is to be able to excel in digital-operational terms. This is possible, because exploiting Big Data is becoming increasingly and rapidly affordable.

With regard to Business Intelligence, *Data Discovery* is the next stage. It helps you combine lucid "magic moments" in your business operations with a significantly better performance via interactive visualization, exploration, planning and execution.

To start with, you can stimulate your power of imagination in inspiration sessions, followed by one or more concretization workshops in which you determine, in conjunction with an organization-wide team, where lucrative Big Data initiatives can best be developed to suit your situation.

The following checklist further concretizes this intention and serves to structure your thoughts and plans and helps you formulate priorities geared to your own situation.

	Your Big Data potential	4	3	2	1
		very much			abso-lutely not
1	Does your organization have strong views on how Big Data could transform your business?				
2	Is the management of Big Data (governance) determined at the top?				
3	Are employees sufficiently involved in the Big Data vision?				
4	Is the collaboration between business and IT such that it can gain maximum advantage from Big Data projects?				
5	Do you have mainly Big Data projects in the areas in which the organization wishes to excel (strategic alignment)?				
6	Does the organization stimulate creativity in order to generate transformative Big Data ideas?				
7	Do you know where to recruit the scarce and elusive data scientists?				
8	If your knowledge workers were to receive the right data more rapidly, would that produce more profit?				
9	Do you possess data that may be interesting to others, and for which there is a market?				
10	Is it important to you to know about purchasing intentions at the earliest possible moment?				
11	Does data philanthropy (donating data to charitable projects) still give you favorable PR?				
12	Have you mastered the regulations for *Privacy by Design*?				
13	Is your Master Data Management in order?				
14	Do you know how to gain access to all your unstructured data?				
15	Have you prepared the infrastructure to cope with converting the further data explosion to concrete business value?				
16	Do you know which Big Data competences you are going to develop, and which you are going to tender out?				
17	Are you really capable of blending external data sources with your own data?				
18	Does your organization have the technology by means of which insights from data can be easily retrieved, instead of having to go searching for them?				
19	Does your organization possess a great quantity of data?				
20	Is there a great deal of diversity in your data?				

References and figures

Capgemini (2012): "18 Reasons Why Your Organization Needs a Chief Data Officer (CDO)," http://www.capgemini.com/blog/capping-it-off/2012/06/18-reasons-why-your-organisation-needs-a-chief-data-officer-cdo

Capgemini (2013): *The Role of the Chief Data Officer in Financial Services*, http://www.capgemini.com/sites/default/files/resource/pdf/the_role_of_the_cdo_in_financial_services.pdf

Capgemini Consulting & MIT Center for Digital Business (2012): *The Digital Advantage: How Digital Leaders Outperform Their Peers in Every Industry*, http://ebooks.capgemini-consulting.com/The-Digital-Advantage

Center for Large Scale Data Systems (2012): *Defining a Taxonomy of Enterprise Data Growth*, http://storage-brain.com/wp-content/uploads/papers/Enterprise-Data-Growth-Ver-1.0-FINAL.pdf

Cisco (2013): *Big Data: Big Potential, Big Priority*, http://newsroom.cisco.com/uk/press-release-content?type=webcontent&articleId=1158061

Cisco IBSG (2012): *Unlocking Value in the Fragmented World of Big Data Analytics: How Information Infomediaries Will Create a New Data Ecosystem*, http://www.cisco.com/web/about/ac79/docs/sp/Information-Infomediaries.pdf

Cooper, M. & P. Mell (2012): "Tackling Big Data," http://csrc.nist.gov/groups/SMA/forum/documents/june2012presentations/fcsm_june2012_cooper_mell.pdf

Daley, R. (2012): "How to Be Successful with Big Data Integration," http://slashdot.org/topic/bi/how-to-be-successful-with-big-data-integration

Fisk, P. (2006): *Marketing Genius*

Franks, B. (2012): *Taming The Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics*

Fraunhofer IAIS (2013): *Innovationspotentialanalyse 2013*, http://www.iais.fraunhofer.de/bigdata-studie.html

IDC & EMC (2012): *The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East*, http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf

IDC & SAS (2011): *Big Data Analytics: Future Architectures, Skills and Roadmaps for the CIO*, http://www.sas.com/resources/asset/BigDataAnalytics-FutureArchitectures-Skills-RoadmapsfortheCIO.pdf

Inmon, W.H. & K. Krishnan (2011): *Building the Unstructured Data Warehouse*

Kalakota, R. (2012): "Organizing for BI, Analytics and Big Data: CoE, Federated or Departmental," http://practicalanalytics.wordpress.com/2012/06/19/organizing-for-bi-analytics-and-big-data/

Krishnan, K. (2012): "Big Data & Analytics," http://www.widama.us/Documents/Big-Data-June-2012.pdf

Krishnan, K. (2013): *Data Warehousing in the Age of Big Data*

Mayer-Schönberger, V. & K. Cukier (2013): *Big Data: A Revolution that Will Transform How We Live, Work, and Think*

Mayer-Schönberger, V. & K. Cukier (2013): "The Rise of Big Data," http://www.foreignaffairs.com/articles/139104/kenneth-neil-cukier-and-viktor-mayer-schoenberger/the-rise-of-big-data

McKinsey Center for Business Technology (2012): *Perspectives on Digital Business*

McKinsey Global Institute (2011): *Big data: The next frontier for innovation, competition and productivity*, http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

MIT Sloan Management Review (2013): "How Starbucks Has Gone Digital," http://sloanreview.mit.edu/article/how-starbucks-has-gone-digital

Moore, G.E. (1965): "Cramming more components onto integrated circuits," http://download.intel.com/museum/Moores_Law/Articles-Press_releases/Gordon_Moore_1965_Article.pdf

Oracle (2012): *Oracle Information Architecture: An Architect's Guide to Big Data*, http://www.oracle.com/technetwork/topics/entarch/articles/oea-big-data-guide-1522052.pdf

Palmisano, S. (2010): "Welcome to the Decade of Smart," http://www.ibm.com/smarterplanet/us/en/events/sustainable_development/12jan2010/

SAS & CMO Council (2013): *Big Data's Biggest Role: Aligning the CMO & CIO – Greater Partnership Drives Enterprise-Wide Customer Centricity*, http://www.cmocouncil.org/download-center.php?id=259

Sattel, G. (2012): "Business Models and the Singularity," http://www.digitaltonto.com/2012/business-models-and-the-singularity

Shirky, C., quoted by K. Kelly (2010): http://www.kk.org/thetechnium/archives/2010/04/the_shirky_prin.php

TCS (2013): *The Emerging Big Returns on Big Data*, http://www.mitcio.com/sites/default/files/sponsorwp/TCS-Big-Data-Global-Trend-Study-2013.pdf

University of Texas (2011): "Measuring the Business Impacts of Effective Data," http://www.wipro.com/images/infographic.jpg

Wal-Mart Labs, http://www.walmartlabs.com/platform

Warden, P. (2011): *Big Data Glossary*

About Sogeti

Sogeti is a leading provider of professional technology services, specializing in Application Management, Infrastructure Management, High-Tech Engineering and Testing. Working closely with its clients, Sogeti enables them to leverage technological innovation and achieve maximum results. Sogeti brings together more than 20,000 professionals in 15 countries and is present in over 100 locations in Europe, the US and India.

About VINT

It is an arduous undertaking to attempt to keep up with all developments in the IT field. State-of-the-art IT opportunities are often very remote from the workings of core business. Sources that provide a deeper understanding, a pragmatic approach, and potential uses for these developments are few and far between. VINT, the Sogeti Trend Lab, provides a meaningful interpretation of the connection between business processes and new developments in IT.

In every VINT publication, a balance is struck between factual description and the intended utilization. VINT uses this approach to inspire organizations to consider and use new technology.

Big Data: The Art of the Possible

The need to chart your Big Data potential is a logical consequence of the purposeful data focus on the combination of business, organization and technology. In this context, you can stimulate your powers of imagination in inspiration sessions followed by one or more concretization workshops in which you determine, in conjunction with an organization-wide team, where lucrative Big Data initiatives can best be developed to suit your situation.

In this fourth and final research report on Big Data, we list the following questions in order to help you formulate a concrete plan. We give a number of reference points by means of which you can draw your own roadmap: the right route for you in the actions you undertake. These questions, just like the checklist at the end of the text, come from our Big Data research reports and from all interaction with the experienced experts with whom we have had contact either personally or online.

Question 1 Why Big Data intelligence?
Question 2 What new insights can I expect?
Question 3 How will these insights help me?
Question 4 What skills do I need?
Question 5 How do Big Data pioneers organize data management and IT processes?
Question 6 How can I merge my structured and unstructured data?
Question 7 Which new technologies should I be watching?
Question 8 What is looming on the horizon?
Question 9 What does this mean in organizational terms?
Question 10 How does this affect everyday life?

You are currently developing the Big Data potential of 2020, and this is why you will find a checklist with twenty questions at the end of this fourth and final VINT Big Data research report. The Report is no simple roadmap from A to B, but a concrete and solid summary of policy recommendations.

You can continue to follow current developments and join in the discussion at http://vint.sogeti.com/bigdata

VINT | Vision • Inspiration • Navigation • Trends