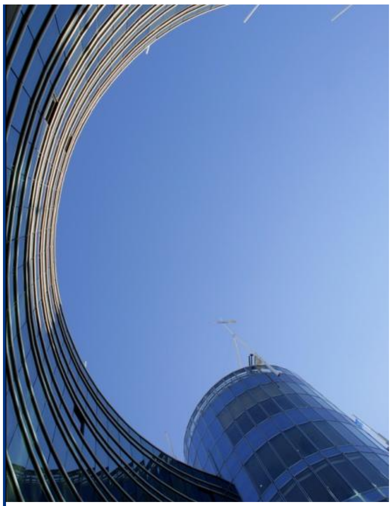
Opinion Paper



Intelligent business by Big Data?

Recommendations for Big Data implementation and improving customer centricity – an approach from a Telco perspective

2013 / 07

Consulting
DETECON

We make ICT strategies work

Table of Contents

1 Executive Summary ... 3
2 Initial Situation ... 4
2.1 Key characteristics of a telecommunications operator ... 4
2.2 A definition of Big Data for telecommunications ... 5
3 The Challenges for Big Data ... 7
3.1 Managing customer perception ... 7
3.2 Dealing with multiple feedback channels ... 8
3.3 Complex customer segmentation and service offerings ... 9
3.4 Organizational issues ... 9
3.5 Shortcomings of legacy systems ... 10
4 The promises of Big Data ... 12
5 Implementing Big Data ... 14
5.1 Knowing which information is valuable ... 14
5.1.1 Analyze the complete value cycle ... 14
5.1.2 Modeling the value cycle taking uncertainties and options into account ... 16
5.1.3 Conduct pilots in order to reduce the uncertainty and update the model ... 17
5.2 Implement a feedback loop ... 18
5.3 Apply Big Data tools sensibly ... 19
5.3.1 Tools for Big Data ... 19
5.3.2 When to use Big Data? ... 21
5.4 Planning for a flexible transformation ... 21
5.4.1 Big Data projects – Success Evaluation ... 24
6 Big Data leads to “Intelligent Business” only when implemented sensibly ... 25
7 Reading on ... 26
8 The Authors ... 28
9 The Company ... 29

1 Executive Summary

Are telecommunication operators able to intelligently improve their business by applying "Big Data"? This paper provides recommendations for telecommunication operators on how to apply Big Data, in order to improve their customer centricity and in the end to guard their customer base. Many of our recommendations also apply to other fields as well.

Telecommunication operators are dealing with customer bases of several millions of clients and have to provide management information for the large organisation with many stakeholders in order to be successful. Most of their revenues are generated by offering services of a complex nature. As a result, data is historically large or let's say "big". Because of the law of "the handicap by a head start[1]", there is conservatism in introducing Big Data.

Most operators serve markets that are saturated. In these markets, retaining customers is a key factor and customer centricity is simply vital. The current hype on Big Data created expectations that Big Data provides additional market relevant information resulting in a competitive advantage for companies that would apply to it.

Yet the environment for making these expectations to come true is challenging. As an example, customer interaction relies on multiple channels, covering a broad spectrum from Twitter™ feeds to written letters and calls. In addition the large organizations are often functionally organized implying separate sales, marketing and technology departments. This often leads to the fact that data is being created in different units and which store their data in different "silos". Consequently, it is challenging to consolidate these silos in order to obtain an integrated view on the actual service situation and therefore the customer.

A key aspect is to capture and to manage how the services provided are perceived by customers. Yet managing this perception is difficult: for example customers take it for granted that telecommunication services are permanently available and therefore complain if services are not provided, despite the fact that service availability being more than 99.9%.

Big Data could indeed offer new solutions, although legacy (structured) mass data has been handled for many years. There is little experience how to handle unstructured data, causing scepticism and conservatism ("What is new about Big Data and "Why do we need it").

In our opinion, Big Data and its analysis will provide a valuable contribution to telecommunication operators and improve customer centricity when some preconditions are met:

- Knowing what information is valuable. (If this is not the case, conduct pilots);
- Implementing a feedback loop across units in order to utilize the gained knowledge;
- Applying Big Data tools in a sensible manner;
- Planning for a flexible transformation.

Only when these essential parts of the value chain are considered and implemented, Big Data can be an intelligent business strategy for telecommunication operators.

[1] http://en.wikipedia.org/wiki/Law_of_the_handicap_of_a_head_start

2 Initial Situation

Large telecommunication operators have to deal with a customer base up to several tens of millions. Almost all their services are based on a complex infrastructure. Therefore, the amount of data handled in telecommunications has traditionally been large, which offers certain challenges when trying to introduce “Big Data” at telecommunication operators.

2.1 Key characteristics of a telecommunications operator

Large telecommunication operators are often “handicapped by a head start” as a result of handling huge quantities of data for many years. This can be easily shown by looking at the amount of operations of a telecommunications operator. For this opinion paper we will use the Deutsche Telekom Group as an example (key figures have been illustrated in figure 2.1).

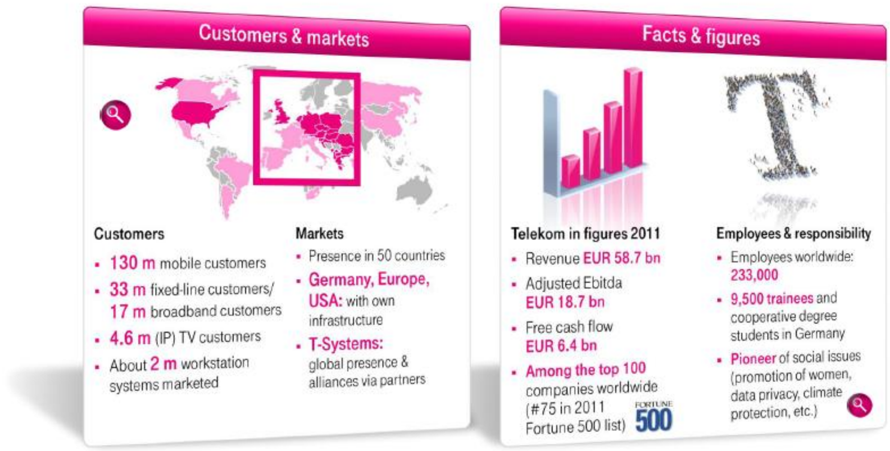


Figure 1: Key figures of the Deutsche Telekom Group

The Deutsche Telekom Group serves nearly 180 million customers on a daily basis. Since telecommunication is an infrastructure service, the customers are in direct contact with daily operations and their mobile phones or DSL routers are always connected to a network. This daily operational contact with customers is a key characteristic of telecommunication service providers. A supplier of a “classical” product like ballpoints has no daily contact with its customers.

In order to enable these daily contacts, telecommunication services rely heavily on IT systems that generate and use considerable amounts of data. For example, each time a customer makes a call, the details of such an event are registered in the IT systems, generating billions of data records with a weekly volume of multiple Terabytes. Even more data is produced across the service chain that is mostly not used by IT yet (e.g. at the network switches).

Not only are the number of customers and the amount of data large, many of the services provided are complex. Telecommunication services have a multitude of options like several data plans or roaming options for example, which make the nature of the service complex and the service portfolio extensive. Moreover, the technical systems needed for making these services feasible are complicated as well.

To manage all these complex operations, DTAG employs 233,000 staff spread over many specialized departments. Many of these departments handle customer care or influence and enhance customer satisfaction. Their success depends on the quality of the information provided by the IT infrastructure out of the massive data flood. These IT systems are therefore key in facilitating and optimizing the analytics on how to serve customers best.

The information required for customer support comes from several data sources:

Production: A considerable amount of data is generated during the production of telecommunication services. For example, for each call a set of meta-information is generated for multiple purposes like network information routing or billing. Such data is highly structured, therefore classical mass data approaches work very well in this area. Each event is processed in the IT systems, resulting in tens of Terabytes of data each week.

Customer service: Customers have multiple options to get in contact with Deutsche Telekom. Various forms like paper mail, email, website contact, call center, shop, SMS and social networks like Twitter™ or Facebook™ are offered to meet the customers' needs. This results in an additional dataflow of several terabytes each week. In addition, this kind of data comes in various data formats and is mostly unstructured. Currently, this data flood is mainly processed and analyzed by humans.

Monitoring: The telecommunications industry provides high availability of service which is permanently monitored by operations. For example server loads and uptime are permanently measured. This kind of data is available in a very structured way.

Let´s summarize: Historically, data has been "large", complex and multi-formatted in telecommunications. In order to grasp the quintessence of the current revolution in IT systems for telecommunication providers, we have to develop a meaningful definition of Big Data that properly distinguishes between legacy "mass data" and "Big Data".

2.2 A definition of Big Data for telecommunications

The open source standard for information management MIKE2.0 (MIKE 2012) provides an useful basis for developing a meaningful definition of Big Data. In this definition, Big Data is not only linked to the volume of the data but also to the complexity, the structure and the number of data sources it is based on.

The Data Warehouse Institute (TDWI) thus defines Big Data with three V's: "Volume, Velocity and Variety" (Russom 2011). Each "V" must be fulfilled in order to classify as "Big Data". Within the current "hype", data is considered to be Big Data already when just one "V" is fulfilled, but we consider the more restrictive TDWI definition, as it has far more explanatory power.

However, the TDWI definition does not consider Big Data in monetary terms. In our opinion, Big Data will only be truly relevant for telecommunications operators when it contributes to corporate targets like increased revenues or cost savings (partly confirmed by Analysis Mason 2013).

We therefore define Big Data as "generating additional **V**alue by analyzing data with **V**olume, **V**ariety and **V**elocity" ("the 4 V's"):

Value is the expected Net Present Value for Big Data investments.

Volume means that Big Data technology should be able to deal with large volumes of data. The promise of Big Data is that multiple tens of terabytes (the size of a current storage solution) can be processed in minutes and a total volume in the range of petabytes (equivalent to size of thousands of current hard drives) can be managed.

Velocity means that Big Data technology should be able to analyze the large volume of data quickly. The performance promised by Big Data systems is the support of right-time data processing. With right-time we mean that the result of data processing must be available when operationally needed, which is not necessarily real-time. For instance, Big Data must be able to interpret the threads on Facebook™ and Twitter™ quickly enough to detect changes in "mood", so that steering customer perception proactively will be possible.

Variety means that Big Data technology should be able to handle several types of data. It should not only be able to find answers in data which is highly structured like transactional data (e.g. billing data), but also be able to create information out of sources with very little structure or none at all, like emails or customer calls.

Up to now, mass data principles are well established in telecommunications for processing billions of data sets of structured data with little variety like call detail records (CDRs). Such data has a format defined by a data model and it is known *what* kind of information is stored *where* it is located in a data set.

Big Data comes into play where mass data has a higher variety in structure (voice messages for example, where it is not known upfront what information comes when), where the number of data sources is high and where the velocity is no longer just batch-oriented. Some applications are:

- *Social Media Analytics:* One popular example for Big Data is social media right-time analytics. Data from different social media platforms mostly contains unstructured text which could be analyzed to detect moods about services;
- *Churn Detection:* In saturated markets, customer churn detection helps to detect dissatisfied customers early. Big Data could enhance the current systems to prevent churn using analytics across all data sources;
- *Predict Network Failures:* One example of machine-to-machine data analytics is the prediction of network failures (Cloudera 2012). Transmission statistics and models could be used to detect patterns before failures occur. These patterns are then detected within the stream of machine-to-machine data allowing proactive fixing of network switches.

3 The Challenges for Big Data

Telecommunication providers are mostly serving mature and therefore saturated markets. Consequently retaining customers is therefore key, whereby customer centricity is essential. That is why we focus on customer centricity in this paper, yet there are many other fields where the application of Big Data within a telecommunication provider could make sense as well. This chapter addresses the challenges for the introduction of Big Data.

Being perceived by customers as *customer centric* is challenging task for a telecommunication operator. Customers take it for granted that telecommunication services function all the time and they complain when services are not provided or fail. When customers complain, they do so via multiple channels covering a broad spectrum from Twitter^{TM} feeds to written letters. Being large organizations, telecommunication operators are often functionally organized with for instance separate sales, marketing, technology and IT departments. This often leads to difficulties to obtain a holistic view on the actual service situation as perceived by the customer.

Big data can be the solution for reaching a consolidated view. However, although legacy mass data (structured data) has been handled for many years, there is little experience with processing unstructured data and on how it can be linked to the classic data world of telecommunications. This causes skepticism and conservatism: “What is really new about in Big Data?”, “Can it really bring solutions to main challenges?

3.1 Managing customer perception

One of the key challenges for a telecommunication operator is being perceived as “customer centric”. One of the reasons for this lies in the nature of the services provided: Telecommunication services have many dissatisfiers [Kotler 2009]. Due to the fact that telecommunication services function 99.9% percent of the time, customers take it for granted that their mobile telephone always works. Given the high availability of the service, there is high customer dissatisfaction on the odd occasion that the service is not available. This is one of the main triggers for customers to complain and to contact a telecommunication operator. The challenge for Big Data is to collect and distribute information in such a manner that this management of customer perception becomes feasible.

3.2 Dealing with multiple feedback channels

If customers would like to contact a telecommunication operator - for instance for a complaint - the channels through which these contacts are established are diverse. They reach from classical paper mail to social media like Twitter™ and Facebook™. The following figure illustrates some key figures regarding customer interaction of the Deutsche Telekom Group.

Figure 2: Key figures of the Deutsche Telekom Group regarding CRM

This wide range of channels brings many complications but also opportunities when applying Big Data.

The majority of customer contacts are calls. So the content (data) within these calls is the most important data source for Customer Relationship Management (CRM). Voice recognition software converts the spoken words into interpretable data. This works reasonably well for voice response systems where the variety of spoken words is mostly limited to 'yes', 'no' and numbers. But a meaningful analysis of customer calls is still a challenge due to the high variety (words, dialects, irony, etc.) and the high volume (millions of calls per year). Currently, most of the calls are analyzed by the call centre agents using classification categories. This means that the main potential data source to be processed for customer centricity remains largely untapped.

The current process of classification is complex and subjective. A call centre agent has to choose out of thousands of reasons to call. Analyses show that each individual call centre agent has its own classification preference. Furthermore, the current classification process is based on historical events: New issues are very often not included in the classification trees and are therefore not captured. This means that it is highly difficult to capture upcoming trends. Moreover, the classification is mostly in very broad categories only, with the result that the real issue the customer is facing is often detected too late.

Voice recognition software and analysis systems develop rapidly, making high volume and high variety text-to-voice conversions possible. Using this technology, Big Data processes could tremendously improve the customer complaint analysis and lead to pro-active trend analysis. Customer complaints would be dealt with more quickly and efficiently leading to improved customer satisfaction. So automated processing and analyzing call centre data seems to be an important use case for Big Data.

Social media like Twitter™ feeds contain similar challenges like calls, since they include the aspects mood, irony and sarcasm as well and sometimes comprise a mixture of several languages. We have tested this in our labs and the results will be discussed in chapter 5.

To implement Customer Relationship Management (CRM) effectively for telecommunication operators, all channels have to be consolidated in order to get a meaningful view of the customer experience and to facilitate customer centricity.

The challenge for Big Data is to consolidate these channels and provide faster and more efficient feedback mechanisms, thus allowing timelier and better reactions to events which impede the customer experience.

3.3 Complex customer segmentation and service offerings

Telecommunication operators tend to target virtually the complete population of a nation as their prospective customers. In order to address the market effectively, many segments with different propositions are defined and sometimes addressed by different brands. Deutsche Telekom in Germany has established two brands: The “Telekom” brand is used for the middle and premium segment, while “Congstar” targets the lower price segment.

Within each brand there are different historical and current tariff schemes, like minute packages, flat rates, pre- and post-paid and many other additional options, leading to a high complexity regarding customers, tariffs, options and products.

As a result, call centre agents have to deal with a very diverse customer base and a complex service portfolio; they must be able to solve multifaceted services issues. To improve customer centricity, the challenge for Big Data is to facilitate call centre agents by offering a more personalized view on the customer and a reduced complexity of the service situation.

3.4 Organizational issues

Most telecommunication operators have long organizational histories with many organizational changes and are often large organizations as well. These aspects lead to further challenges for implementing Big Data.

Until the early 1990s, most operators focused on fixed network services. The large scale introduction of mobile telephony resulted for most operators in the founding of a separate organization that focused exclusively on mobile telephony. Around the year 2000, the internet boom started and most operators founded yet another separate unit focused on online services. After the boom expired, the consolidation of the three separate organizations started. Currently, within most operators one organization per geographic area offers a complete portfolio of telecommunication services, including cloud services and TV.

All these organizational changes are also reflected into the IT infrastructure: The initial segmentations and subsequent consolidations lead to large organizations with a functional division of work and long process chains. Within every step of these chains, a specialized view on the data was required. This leads to local optimization at each link of the chain but the overall view on the customer was neglected. Many legacy systems from previous organizational structures are still being used, making it complicated to obtain a complete view on customer perception.

In summary: Most telecommunication operators have no centralized view on the fixed line, mobile and IP services and the underlying processes in order to manage customer perception holistically. This challenge of cross domain and cross company analysis can hopefully be addressed when introducing Big Data.

3.5 Shortcomings of legacy systems

Introducing Big Data in a telecommunication operator is challenging because its IT landscape is not a “green field situation”. As mentioned previously, a high volume of data is processed in the legacy / structured data world.

The next picture illustrates a common way data within an organization is handled today. For simplification purposes, this example focuses on the business support systems (BSS) for mobile services only. A similar picture also exists for the network and is called the operational support systems (OSS). Moreover, the fixed line and the IP services are supported by similar BSS and OSS architectures.

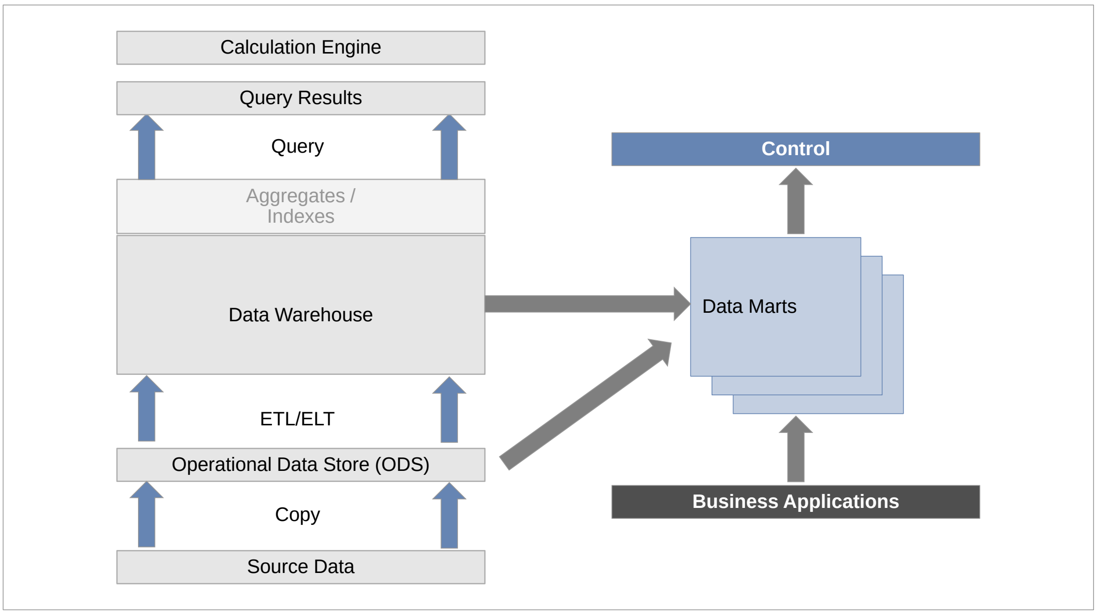


Figure 3: The current way of handling data (Elliot 2011)

As seen in Figure 3, many sources feed into a Data Warehouse (DWH). The DWH itself is the source for so called “Data Marts”. Data marts are designed to provide data for specific purposes, e.g. one data mart for business analytics (finding of patterns, score calculations, etc.), one for campaign management (covering all data related to make offers to customers), one for finance (covering risk aspects) and many more. For each data mart, the data is prepared, aggregated and filtered so that they fit the purpose of the data mart best. But there is no single place where all data is gathered and used in a holistic way. Further complicating the process is the synchronization of the existing warehouses.

So for example, just within the business related data marts, a common view on a customer does not exist. Customer care data for example is not properly linked to campaign data and risk data not connected to sales data. Many organizations work hard to improve this situation, they aim to have a “360° view” about the customer. But even when all the different sources of customer data are merged, data originating from the network, social media and machine-2-machine communications is not included. Each piece of information is either stored in its own silo or even completely ignored.

There are three main reasons for not using all available data: First of all, data 'belongs' to specific departments within the organization and there is no overall data stewardship; there is no central role for a dedicated accountability of all data. The second reason is that the current IT landscape of most telecommunication operators is not mature and cheap enough to process all data. Last, data security and regulatory requirements impose restrictions as well.

Adding new types of data into your analytics implies new processes and new ways how to think about data – so a change of mind is also required. Just continuing with legacy systems is no option to survive in future. But adding Big Data to the existing IT landscape will be a challenging task. Whether this challenge is worth the effort depends on the promises Big Data can deliver.

4 The promises of Big Data

For decades, telecommunication business units have developed specific CRM strategies including data mining, predictive modeling and further analytical concepts, to get a deeper insight of the customers wishes and to treat them accordingly. One paradigm used was 'The more information you have the better you can treat your customers'. The master in this discipline was and still is Google, “trying to collect any available information” (Google, 2012). But the trade-off may be that a wealth of information leads to a poverty of attention (Simon, 1971).

That is why any CRM strategy is focusing as much as feasible on the customer in order to deliver an optimum service quality. The approach taken is to interact with the customer via the right channel, at the right time of the day in the right frequency, while providing the right content. The data used for that comes from billing, mediation, CRM, ERP and other sources. All is established and practiced over years. So where is the change now? What are the promises of Big Data? Does it solve the challenges?

Coming back to the four V's, “generating additional **V**alue by analyzing data with **V**olume, **V**ariety and **V**elocity ", we will make some illustrations of this. In the picture below, a “classic” CRM strategy is displayed on the left side. The consequences of implementing Big Data are displayed on the right side.

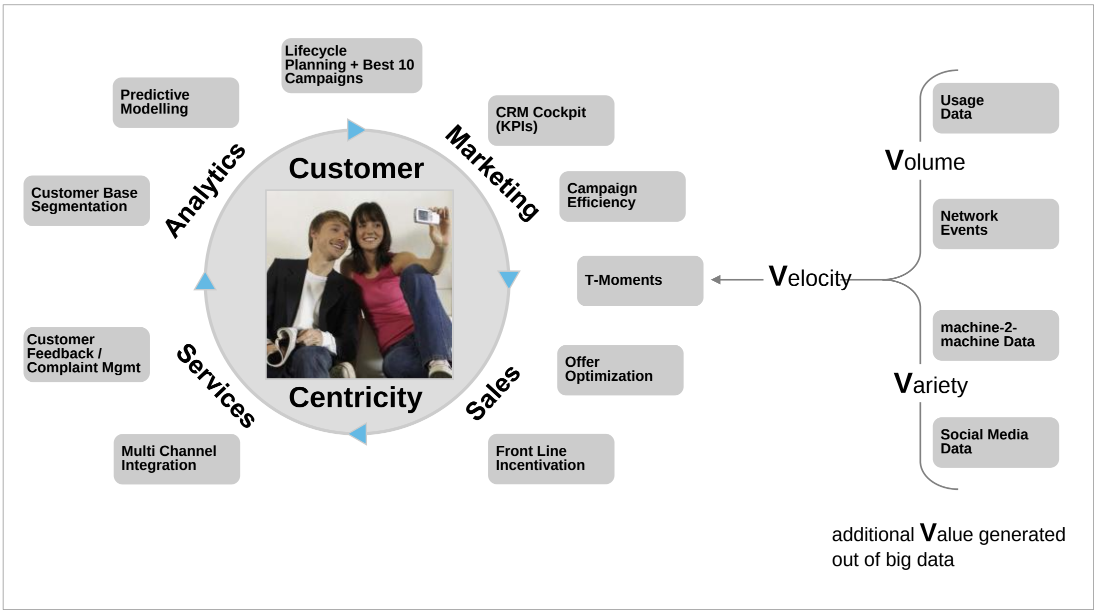


Figure 4: Customer centricity by big Data

With Big Data:

More data with high **Volume** (e.g. network events) and high **Variety** (e.g. unstructured data from social media) is

linked at the right time (high **Velocity**) with the CRM world, to trigger interactions with the customers with the right content via the right channel at the right time,

to generate additional **Value**, for example by improving customer satisfaction, reducing churn or selling additional goods.

One idea for improving customer satisfaction and efficiency simultaneously is the linking of network load with the number of customers using mobile services in a region. In case of public events like concerts, the resulting change in customer number in a region can lead to network overload and thus bad customer experience. To prevent this, additional mobile communication equipment is already used for well-known large events like the Oktoberfest in Munich (Teltarif 2012).

But smaller and spontaneous events are not covered. By linking the network capacity data and the customer impact in right-time, new ways for dynamic network capacity management can be realized to reach the best customer experience with most efficient and economic use of resources. However, integrating Big Data into CRM strategies is not clear cut and can be challenging.

For example, why analyze social media when we can ask agencies to execute customer surveys and to evaluate the acceptance of our products? The answer is that although complexity is added when introducing Big Data, at the end automation and simplification will be the result. Bearing this in mind, promising implementations for Big Data are:

- Automated triggering of customer care agents, to join threads on Facebook™ when strong bad experiences surface;
- Automated triggering of technical services, when operational data indicate that a switch will fail soon; automated prioritization based on the number of customers impacted by a failure;
- Automated finding of keywords and thus the current trends used by customers when calling the customer care (across the thousands of call centre agents).

This data was not used in the past, mainly due to the fact that the technology was not ready to process that amount of data economically. The technology has matured now, yet the prospective business value of Big Data is not clear upfront.

In other words, Big Data means having to explore to find the value. It is like “finding the proverbial needle in the haystack” that enables additional value. You have to invest in technology, you have to invest in data exploration and you have to accept failures, meaning the risk of no value out of some data or no immediate return of invests.

But at the end of the journey you have capabilities that enable additional revenue, less churn and differentiation to competition. That is the promise of the four V's.

But does it solve the challenges indentified? Can Big Data improve the management of customer perception, consolidate customer feed-back over many different channels, support the call centre agent with personalized view on the customer, unify information over different domains and integrate with legacy systems? According to our opinion this is only feasible if big data projects are handled in the right manner.

In the next chapter we describe what needs to be done to materialize these promises.

5 Implementing Big Data

The implementation of Big Data within a telecommunication operator is not easy at all. The promises made are difficult to implement; in order to introduce Big Data effectively, many points need to be addressed. Just buying the machinery and hoping for success is not the right strategy for sure.

We think that Big Data analysis can provide a valuable contribution to telecommunication firms when - besides taking care of data security and privacy - these conditions are met:

- Knowing what information is valuable. (If this is not the case, conduct pilots);
- Implementing a feedback loop across units in order to utilize the gained knowledge;
- Applying Big Data tools in a sensible manner;
- Planning for a flexible transformation.

5.1 Knowing which information is valuable

In the current „hyped“ situation, Big Data is supposed to result in value instantly by just buying it. We do not think this is the case at all. In this paragraph we focus on the methodology we think is appropriate to determine which information is valuable. The methodology comprises the following elements:

1. Analyze the complete value cycle;
2. Model the cycle while taking uncertainties and options into account;
3. Conduct pilots in order to reduce the uncertainty and update the model.

5.1.1 Analyze the complete value cycle

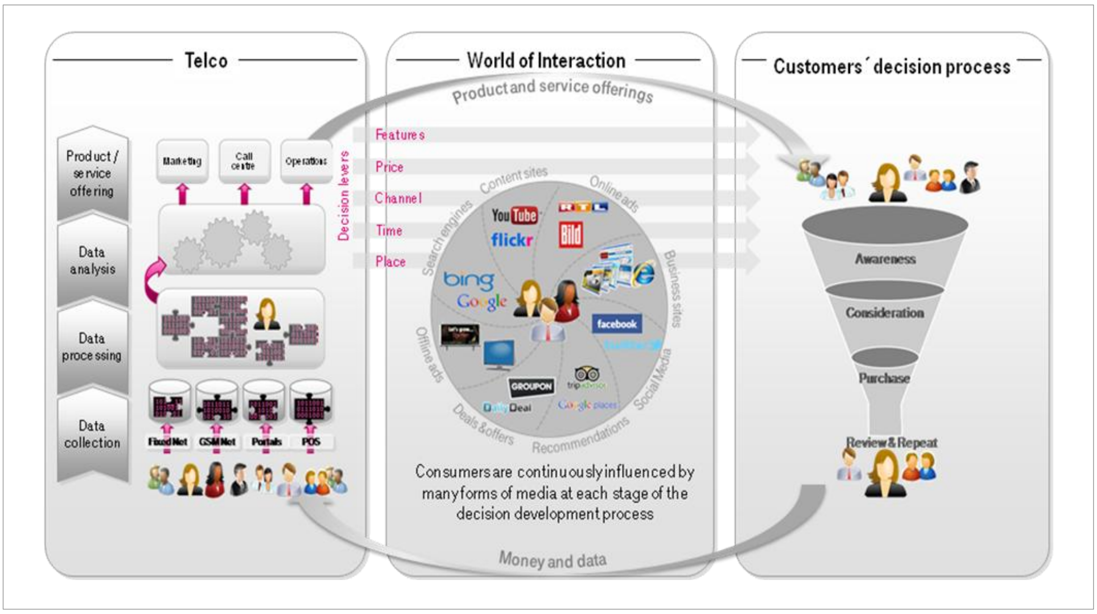


Figure 5: The value cycle

In order to define which information is valuable, the whole value cycle needs to be considered. Figure 5 depicts this value cycle. The current focus of Big Data is mostly on the left side of the picture. This side depicts the supply side of the cycle, consisting of the telecommunication operator with its IT landscape, the information process chain and the service supply side of the cycle.

The right side of the picture represents the demand side, where the customer decision processes have to be influenced. In order to influence the customer favorably, the telecommunication operator has to focus on the demands of the customer and behave in a customer centric manner on the market.

The middle of this cycle, the “world of interaction”, represents the market where supply and demand meets and product and service offerings are exchanged for money and data.

Analyzing the value cycle’s many parameters can only be determined with a high amount of uncertainty. It is our hope that Big Data can decrease this uncertainty, also for determining if an investment in Big Data is economically viable. Consequently there is a “chicken and egg problem”: Whether Big Data is economically viable can only be ascertained by Big Data analysis. To break this vicious circle, pilots with Big Data could offer a way out.

When assessing the economic benefits of Big Data pilots, “classical” net present value methods are inadequate: Modeling the cycle is complex and many parameters can only be determined with a high amount of uncertainty. This uncertainty has to be taken into account. Second, classical net present value calculations use a static approach, neglecting all kinds of options which might occur during the lifetime of the asset. These options should be taken into consideration for a truly comprehensive and meaningful calculation of the economic benefits.

In the next paragraph we will explain the methodology needed for these kinds of evaluations.

5.1.2 Modeling the value cycle taking uncertainties and options into account

The problem of high uncertainty can be solved by modeling uncertainty explicitly and applying Monte Carlo simulations to calculations. The second problem of how to consider and evaluate different options before making a final investment decision, is solved by conducting pilots and taking several options into account in the economic evaluation. We will first discuss how to model uncertainty.

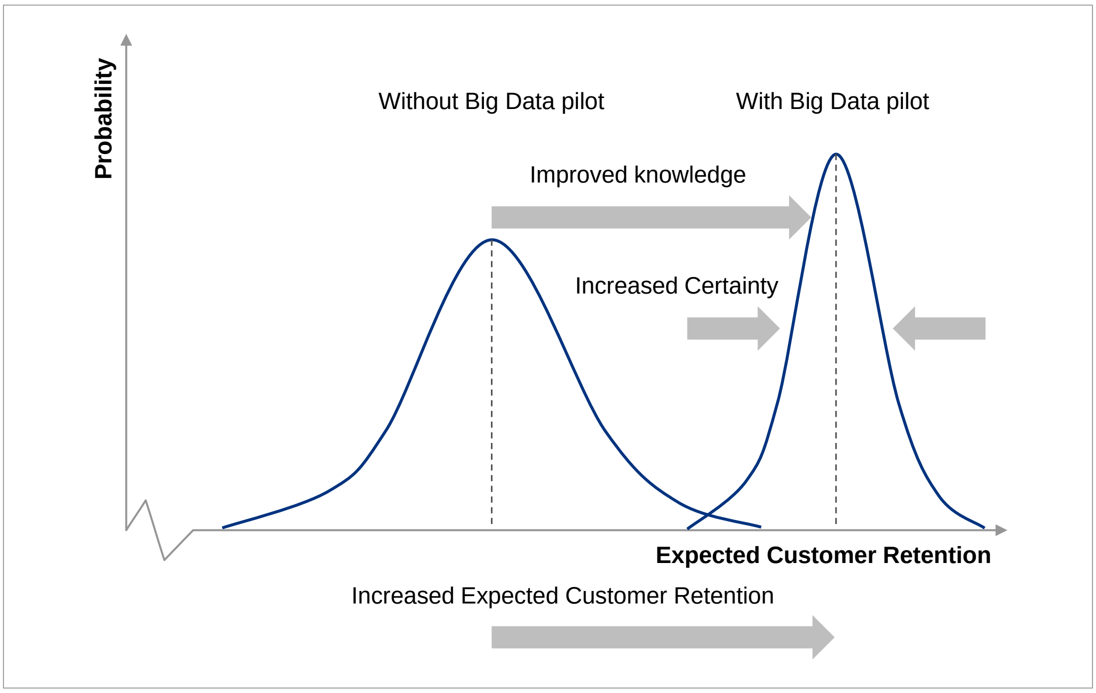


Figure 6: Modeling uncertainty and the influence of pilots

Figure 6 depicts an example for modeling uncertainty. The parameter “Expected Customer Retention” is plotted as a probability distribution. By introducing a probability distribution the uncertainty can be captured. Without conducting a pilot, the expected customer retention when introducing Big Data can only be estimated with high levels of uncertainty, depicted by a wide curve on the left side of the picture. In case of a successful pilot with Big Data, the knowledge about the customer increases; this is depicted in the right side of the picture. The assumption is that through the increase of knowledge about the customer, the resulting customer retention increases (shift to the right). At the same time the uncertainty decreases, which means the probability distribution gets smaller.

With this type of modeling the most promising use cases can be identified, which subsequently have to be tested in a pilot.

5.1.3 Conduct pilots in order to reduce the uncertainty and update the model

As described, we face various uncertainties in Big Data projects and they will have a massive impact on the whole organization. The risks brought on by implementing a new solution and the according processes should be therefore minimized.

We do recommend using pilots of Big Data in close cooperation with suppliers for risk reduction and involving all the relevant stakeholders within the firm in them.

Some of the reasons for pilots are the following:

- They result in real information what Big Data could entail for a particular telecommunication operator and therefore reduce uncertainty;
- Pilots shorten the learning curve of Big Data implementation, thus improving quality of the decision process;
- The cooperation between the suppliers and the telecommunication operator in question is assessed as well.

The pilots should be conducted based on agile methodology, yet with very clear success criteria to be defined up front. By conducting pilots, essential issues can be assessed:

- Big Data comes with many tools (details will be covered in chapter 5.3.) based on the so-called Hadoop framework. Pilots support the evaluation and selection of appropriate tools for the Big Data use cases;
- Finding of limitations of Hadoop-based tools, e.g. when integrating with Facebook™;
- Solving challenges when working on text mining in a multilingual environment.

What we do not recommend is sticking to the traditional “waterfall” method and starting with a paper based approach, spending huge sums of money on studies. A pilot is probably less expensive. In our opinion, paper based exercises for selecting Big Data only lead to speculation of what Big Data could mean, as well as to inflated business cases and therefore poor decision making.

On completion of the pilots, evaluations can be conducted and the Monte-Carlo models can be updated. If there is a positive business case, a step wise approach is recommended.

All required departments need to be involved in implementing the solution for production. A production team will be formed and a classical solution roll-out process will start. While the team size of previous phases should not exceed 5 to 10 people, the team should be increased at this point in time to be able to deliver the solution.

5.2 Implement a feedback loop

In order to gain benefits from the advanced analytical capacities of Big Data, the Velocity (the 4th "V") and the possibilities for proactively engaging the customers, processes and systems need to be adapted. Switching from a reactive to a proactive mode also requires the adaptation of the working methods and scripts and the implementation of a comprehensive feedback loop.

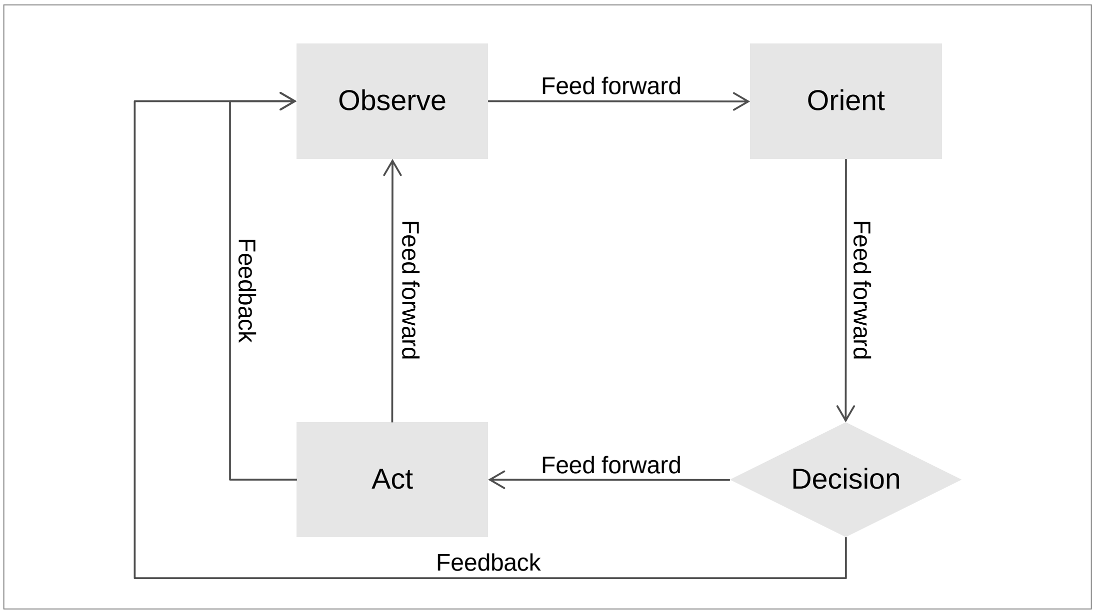


Figure 7: Simplified OODA-Loop by John Boyd (Klausnitzer 2013)

The model of OODA (Observe, Orient, Decide and Act) -Loop was introduced by John Boyd (Figure 7). He analyzed the decision process in a fighter jet during a fighting decision and found out that the main parts in this process are gathering information (Observe), processing these information (Orient), decisions about actions (Decide) and the action itself (Action).

The collected information is used in a feedback-loop in order to analyze and optimize the decision process of future actions as a constant learning and optimization process. The OODA-Loop can perfectly be adapted to describe a comprehensive feedback-loop for the usage of Big Data in enterprises.

The first step is the collection of structured data from internal sources and systems as well as probably unstructured or poly-structured data from sources like social media channels or call-center calls. While processing the data, it is key to figure out for which department the information can be useful. If we think about telecommunication companies there can be information about customer needs, which is valuable for the marketing or research and development. But there can also be data about problems with the firm's services and products which are very useful for the product development and the service departments. In case of problems with products or services, the decision can be evaluated either to contact the customers proactively in order to preempt the problem or to proof that the problem had not come up.

After a decision is made, e.g. the campaign will be executed, the feedback loop is closed with the information out of the actions. This information is very valuable for learning and optimizing campaign processes, service processes or the development of new products and services.

There are six major factors that influence an effective Big Data feedback loop (Klausnitzer 2013):

- The quality and quantity of data;
- The quality and speed of the data transformation to relevant information in order to provide a suitable basis for decision-making by appropriate algorithms;
- The quality and speed of action-taking;
- The quality and quantity of the data gathered from activities;
- The quality and speed of the transformation of this data to relevant information in order to provide a suitable basis for decision-making, by appropriate algorithms for feedback generation;
- The quality and speed of bringing this feedback information back into the feedback loop.

5.3 Apply Big Data tools sensibly

Within many Big Data discussions, the term Big Data is often equated with Hadoop. We see it differently (see 2.2). For us, Hadoop is just one of the technologies to support the 4 V's of Big Data and thus to handle high Velocity, high Volume and high Variety to deliver additional Value. The application landscape is much wider and one has to distinguish between various open source and proprietary products.

Because of the Brownfield situation, both tool selection and deciding for which situation the Big Data platform will be applied need careful consideration.

5.3.1 Tools for Big Data

We will not “deep dive” into the technical details about Hadoop or any of the other tools. But let us clarify some aspects: Hadoop is based on the Apache Hadoop Project (Apache 2012), an open-source framework for handling data-intensive processes on a distributed environment based on papers by Google. Much of current discussions focus on the Hadoop Distributed File System (HDFS), currently assumed to be the cheapest way to store data on a big scale, but it is only a part of the entire suite. Many other tools are required in order to create a functioning Big Data environment.

There are many difficulties when implementing existing open source solutions for Big Data like Hadoop, Drill or Storm. To implement open source solutions, the core technology needs to be understood, use cases to be developed and maintained. Various companies offer development and support for custom Big Data solutions which offer value.

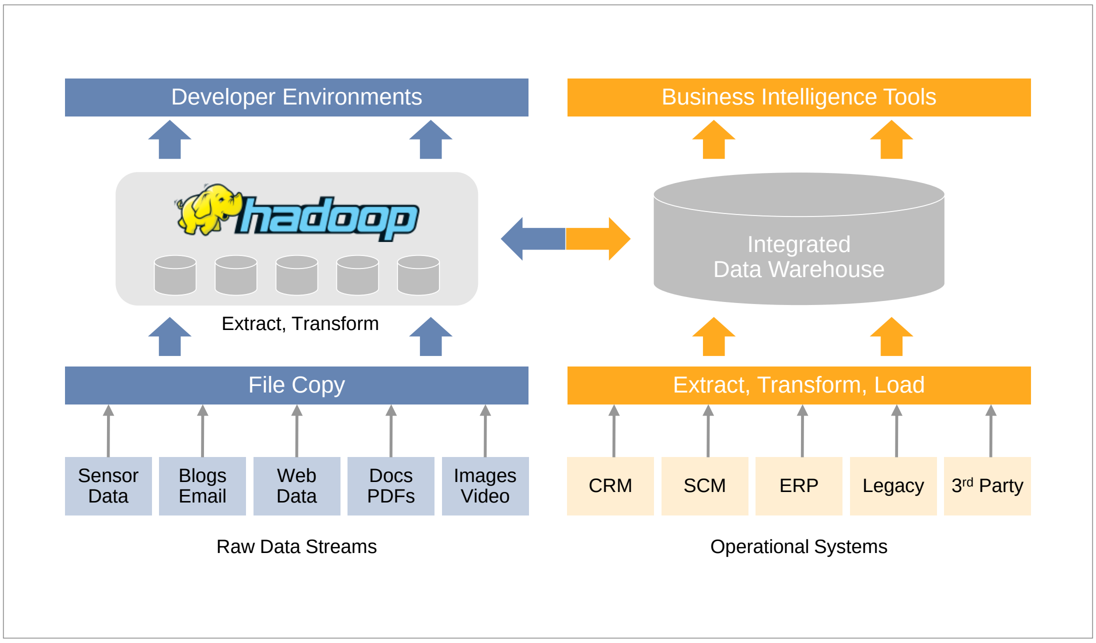


Figure 8: Big Data next to the EWH (Awadallah et. al. 2011)

It is expected that tools developed for Big Data are moving into the classic world of data warehousing (Kobielus 2011 and IBM 2012a). This will combine the strengths of Big Data in capturing and storing of data with the strengths of data warehouses (DWH) in analysis and presentation (Dumitru 2011). However, the DWH itself, even with Hadoop technology, for us does not belong to Big Data.

Rationales for Big Data

- The analysis of high volume combined with high variety data, for instance unstructured data
- Analytical work on raw data (e.g. sand boxing via Apache Mahout)
- Exploration of data that do not have to be archived (analyse and forget, e.g. finding of patterns that are used for complex event processing)
- Cost aspects (capturing and storing of high volume data)
- High velocity but no transaction safety needed in data processing
- Storing of data that might be useful in the future (parking position for data)

Rationales for classical Data Warehouse

- Structured data only, eventually in high volume
- Data to be archived for a longer period
- Data to be historized (changes of data over time stored)
- Needs for data quality (validity of data, transaction safety)
- Data to be included in Meta Data Management
- Data for corporate reporting
- Data for transactional processing
- Data to be enriched, integrated or cleaned with other data sets

Figure 9: Rationales for use cases belonging to either Big Data or DWH

Within our BI architecture, we see Big Data as an additional box next to the DWH. This fits to the way vendors see this topic, as illustrated in the next picture (Awadallah et. al. 2011) by placing a Big Data box (here called Hadoop) next to a DWH.

5.3.2 When to use Big Data?

With Big Data becoming a new box within the IT landscape, the question is when and how to make use of this box. As always in life the answer is not black or white. There are areas that clearly speak for Big Data while other areas require a DWH. The following overview lists criteria for deciding whether a use case belongs to Big Data or the "classic world" of DWH.

Figure 5.5 shows that the criteria for using Big Data are strongly related to either unstructured data or using a cost efficient way for capturing and storing of data that is used for one-time analytics. The more aspects like data quality, integration to other data or historization comes into place the more the use case belongs to data warehousing.

Big Data enables analytical evaluations that are either not possible or not economically within the current DWH landscapes. The latter point is seen also on vendor side. For example, SAS (one of the world's biggest companies for analytical solutions) offers high-performance analytics for structured data in both worlds, thereby on the Big Data side solely the Hadoop distributed File System is used.

Below some examples on how to determine whether a use case belongs to Big Data or DWH are listed:

- Text mining, e.g. of social media feeds: High Volume: yes; high Variety: yes (belongs to unstructured data), high Velocity: yes → Big Data;
- Store as less as possible in DWH philosophy, i.e. do not load the maybe needed data into DWH: High Volume: yes, high Variety: no, high velocity: no, but cost aspects: yes → use Big Data;
- Store for CDR analysis, i.e. place for analytics on all call detailed record data: High Volume: yes, high Variety: no, high Velocity: no, in case fire and forget is possible → Big Data, in case data needed for e.g. customer complaints → DWH;
- Click stream analytics: High Volume: yes, high Variety: no (clicks have structure), high Velocity: yes, do I just want to analyze the clicks → Big Data; do I want to link the clicks to customers and campaigns → Big Data for the analysis, results to be loaded into DWH, used there for advanced campaigning

5.4 Planning for a flexible transformation

Recent TDWI surveys (Russom 2011, 2012) show that Big Data currently is a hype topic. The surveys also highlight that not having the appropriate technology and in-house skills is one of the worries of senior IT management today. Vendors' tools though do not automatically ensure concrete business results.

There are many papers on performance increases in analytics and all the calculations after applying new technologies (e.g. SAS 2012b, Huang 2012) and the results are tremendous. But one thing is not mentioned namely the concrete marketing lift: How much more was sold via this Big Data activity?

This “lift” criterion is one of the most crucial numbers in marketing. For decades, expertise has been build up on how to define appropriate control groups and twin pairs of customers with same attributes like revenue and churn probability, with the results of all marketing activities then being correlated between target groups and control groups to measure the effectiveness and the lift of the underlying campaigns.

But Big Data is new. For example, when analyzing and reacting on machine2machine data there are no control groups yet and therefore, the business success is hard to measure. It is not even proven that each and every activity will have positive business impact. In other words: Introducing new topics like Big Data for high-performance analytics requires the acceptance of failure and delay and thus potential project cost overruns (Soong 2012).

With Big Data, we are at the beginning of a journey. Still we are in the exploration and discovery phase. It offers a chance to differentiate from competition and being a first mover with new insights that lead to competitive advantages. But the return on invest and the business case cannot be stated upfront and flexibility has to be build-in in the transformation plan. Accepting this is a must before starting a Big Data implementation project.

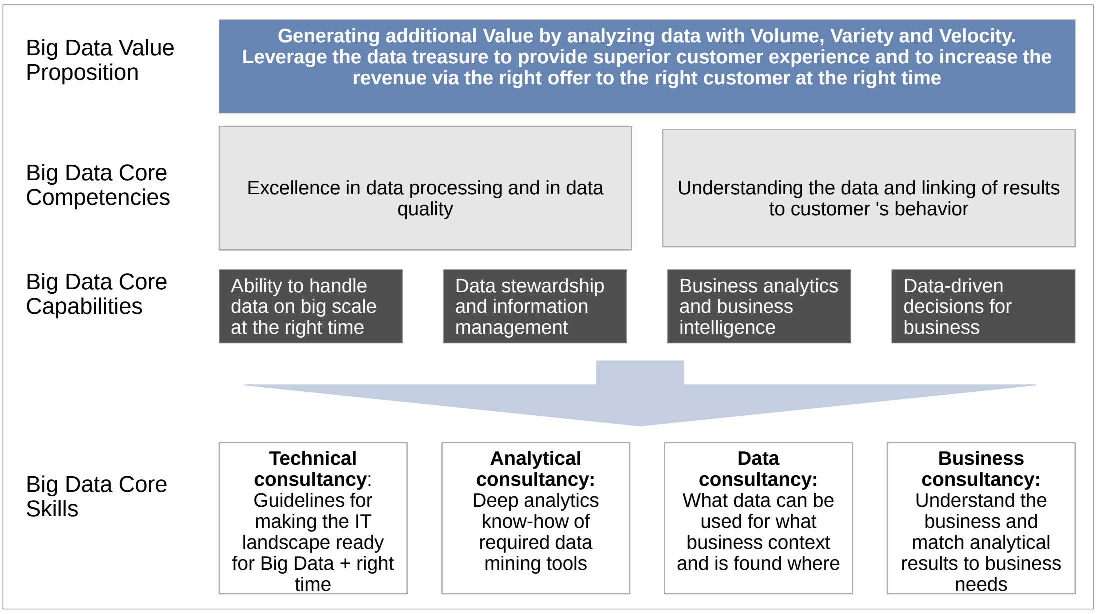
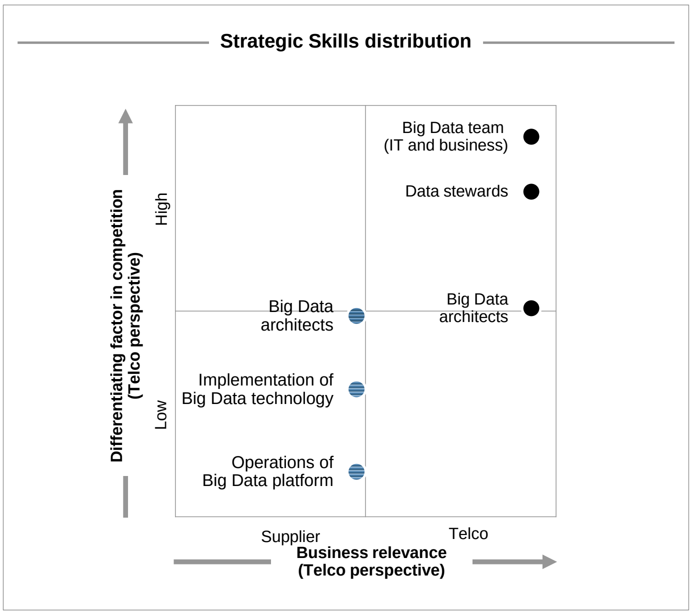


Figure 10: Value creation map for Big Data

As seen in figure 10, four skills are essentially needed for Big Data projects:

- Technical consultancy: Prepare, build and operate the IT landscape for Big Data;
- Data consultancy (also called 'data stewardship'): Knowledge about available data and where to find it;
- Analytical consultancy: Analytical know-how to find patterns within the data (data exploration);
- Business consultancy: Match the business with the analytical results to find the treasure (translating the results from highly academic statistics and algorithms to the business world and deriving of concrete business tasks out of those results).

A sourcing strategy for these skills is shown within the next figure.



Sourcing strategy

- In-house:
  - Big Data team (analytical and business consultancy): Analytical experts to find patterns combined with business-IT-experts who are able to talk business and to discoverer the value out of the data
  - Data consultancy/stewards Master of the data
- Candidate for outsourcing:
  - Technical consultancy like
    - Design and setting up an architecture for Big Data
    - Implementation of IT infrastructure for handling Big Data
  - Operations of IT infrastructure for Big Data

Figure 11: Skill matrix and sourcing strategy for Big Data

According to Figure 11, a potential sourcing strategy for Big Data is to have the analytical and business expertise in-house while the whole implementation project starting from the design, via the system architecture, the deployment and finally the operations can be out-sourced. Even the complete project management for a Big Data implementation project can be outsourced.

Currently, able experts to run the required statistical analysis and to translate the results into business functionalities are rare. Therefore, it is an option to even source the analytical experts from the external market. But the translation into business functionalities and the definition of the resulting business activities must be under in-house control, since the differentiating factor of Big Data lies exactly here.

5.4.1 Big Data projects – Success Evaluation

One weakness of Big Data is lacking of concrete business results. Therefore, it is important to define measurable targets at the beginning of a Big Data project. The targets strongly depend on the use case. The following examples indicate how measurable targets can be established:

- Use case focus on improved data analytics:
  - List all new capabilities;
  - Overview of data and volume that is analysed now (what was not analysed before);
  - Results and findings.
- Use case to derive new insights that shall be used to start campaigns:
  - Describe how the change pays into reaching a higher maturity in campaigning;
  - Acceptance rate of the campaigns;
  - Revenue of the campaigns.
- Use case pays into customer experience:
  - Customer surveys before and after activity;
  - Number of customer interactions initiated by Big Data project.

Best practice is to not underestimate the effort needed to make the results of Big Data tangible. That is why benefit management should be started at the beginning of each Big Data project.

6 Big Data leads to “Intelligent Business” only when implemented sensibly

Big Data is currently a “hype” in “telecommunication land”. To see clearly what is good about it and what is not, we first suggest defining what Big Data is. In our opinion, the best definition is technology agnostic: “Generating additional **V**alue by analyzing data with **V**olume, **V**ariety and **V**elocity” (“the 4 V's”)

The possible benefits of Big Data for analyzing high volume calls and social media suggest that it instantly creates value for CRM, solving major challenges in CRM like consolidating customer feed-back over many channels, supporting the call centre agents with personalized views on the customer and providing unified information over different domains. In our opinion, these benefits are realistic possibilities but only when Big Data projects are handled in the right manner.

We think that Big Data analysis can provide a valuable contribution to telecommunication firms when - besides taking care of data security and privacy - these conditions are met:

- Knowing what information is valuable. (If this is not the case, conduct pilots);
- Implementing a feedback loop across units in order to utilize the gained knowledge;
- Applying Big Data tools in a sensible manner;
- Planning for a flexible transformation.

Only when these essential parts of the value chain are considered and implemented, Big Data can be an intelligent business strategy for telecommunication operators. When introducing Big Data, it is very difficult to know upfront which use cases results in valuable information and a positive business case. The value cycle and Monte Carlo simulations are valuable tools for analyzing where benefits might accrue and to define sensible use cases. Yet pilots are required to feed the modeling with valuable information and to learn about Big Data in order to make sensible business decisions.

In order to profit from the advanced analytical capacities of Big Data, the Velocity (the 4th “V”) and the possibilities for proactively engaging the customers, processes and systems need to be adapted. Switching from a reactive into a proactive mode also requires the adaptation of working methods and scripts and the implementation of a comprehensive feedback loop. Big Data has to be implemented in a Brownfield situation: Big Data cannot replace the existing legacy systems, both will have to coexist.

The introduction of Big Data requires a flexible transformation with new working floor skills which are hard to find on the labor market and which have to be learned. Decisions have to be taken which skills must be kept in-house and which can be outsourced. In this matter we think that technical consultancy and operations are the only potential candidates for outsourcing. The targets need to be defined upfront but the implementation should be driven by agile projects. Big Data is new and opens interesting opportunities. Introducing Big Data admittedly reveals a learning curve, yet by implementing it sensibly and systematically, telecommunication operators can reap great benefits from the latest developments in technology and information management: Intelligent Business through Big Data.

7 Reading on

Analysis Mason (2013), Patrick Kelly and Anil Rao, *Big data analytics: how to generate revenue and customer loyalty using real-time network data*, Analysis Mason 2013

Apache (2012) *Hadoop*, found on 27-SEP-2012 under http://hadoop.apache.org/

Awadallah, A., Graham, G. (2011) *Hadoop and the Data Warehouse: When to Use Which?* Published 2011 by Cloudera and Teradata

Copeland, T, Antikarov,V (2003), *Real Options a practitioner´s Guide*, Thomson-Etexere

Cloudera (2012) *Ten Common Hadoopable Problems*, Whitepaper, found on Cloudera homepage on 22-AUG-201

Dumitru, A. (2011) *Hadoop and the Data Warehouse: When to Use Which?*, found on 29-AUG-2012 under http://www.youtube.com/watch?v=IjpU0dLIRDI published 27-JUN-2011

Elliot, T. (2011), *Why the last decade of BI best-practice architecture is rapidly becomming obsolete*, found under http://timoelliott.com/blog/2011/03/why-the-last-decade-of-bi-best-practice-architecture-is-rapidly-becoming-obsolete.html, posted 03-MAR-2011

Google (2012), *Google's philosophy*, found on 26-OCT-2012 under http://www.google.com/about/company/philosophy/

Huang, P. (2012) *Predictive Analytics Automation: Make better decisions faster*, presentation at Teradata Partners Conference 20-24-OCT-2012

IBM (2012) *Hadoop: Nucleus of the Next-Generation Big Data Warehouse*, found on 29-AUG-2012under http://ibmdatamag.com/2012/08/hadoop-and-data-warehousing/ published 06-AUG-2012

Klausnitzer, R. (2013); *Das Ende des Zufalls: Wie Big Data uns und unser Leben vorhersagbar macht,* Ecowin Verlag (2013)

Kobelius, J.-G. (2011) *Hadoop: Future Of Enterprise Data Warehousing? Are You Kidding?*, found on 29-AUG-2012under http://blogs.forrester.com/print/james_kobielus/11-06-08-hadoop_future_of_enterprise_data_warehousing_are_you_kidding; dated 08-JUN.-2011

Kotler, P. (2009), *Marketing Management*, Pearson Education, 2009

Microsoft (2012) *Strategies on Big Data*, promotion paper published by Microsoft Hellas in 2012 edited by Nikos Makris

MIKE (2012) *Big Data Definition*, found under http://mike2.openmethodology.org/wiki/Big_Data_Definition on 27-SEP-2012

Russom, P. (2011) *Big Data Analytics*, TDWI Best Practice Report Forth Quarter 2011, TDWI Research, Renton, USA, originally developed in TDWI blog posts available under tdwi.org/blogs/philip-russom

Russom, P. (2012) *High-Performance Data Warehousing*, TDWI Best Practice Report Forth Quarter 2012, found on 28-SEP-2012 under http://tdwi.org/research/2012/10/tdwi-best-practices-report-high-performance-data-warehousing

SAS (2012a) *In-memory Analytics for Big Data*, SAS Whitepaper, no publication data, found on www.sas.com in August 2012

SAS (2012b) *SAS Roadshow Big Data and High-Performance Analytics*, held in Cologne, Germany on 03-July-2012

Simon, H. (1971) *Computers, Communications and the Public Interest*, pages 40-41, Martin Greenberger, ed., The Johns Hopkins Press, 1971.

Soong, T. M. (2012) Quote: "*Board members must be willing to accept new things and to accept failures. So budget for failure is needed*.", best practice report, given in a face2face at the SAS Premium Business Leadership Series, 08-09-May-2012

Teltarif (2012) *Start des Münchener Oktoberfests: Hier glühen nicht nur die Köpfe - Mobilfunker wollen immense Netzbelastung durch Aufrüstung stemmen*, published 22-SEP-2012

8 The Authors

Frank Wisselink gained a wide experience in industry reaching from oil and gas sector, electronics to telecommunications before joining Detecon. The last 6 years Frank managed and contributed to key innovation, change and strategy projects within and outside the Deutsche Telekom Group. He specialises in optimising the business benefits of innovations like Big Data by integrating strategical, marketing, financial and organisational aspects. Frank holds a PhD in engineering and a BSc (Hons.) in Economics and Management of the University of London International Programm conducted by the London School of Economics and Political Science.

Frank can be reached at +49 22191612946 or frank.wisselink@detecon.com

Lars Hofmann works as Lead Designer Business Intelligence at T-Systems International GmbH. Within the Group Design Competence Centre he is accountable for innovation in BI and advanced analytics. Currently he focuses on the end2end-integration of new BI capabilities as well as simplification of legacy BI landscapes. His broad experiences cover the domains billing, payment and business intelligence in telecommunication. Lars Hofmann holds a doctorate in physics and a M.Sc. in Enterprise Information Management.

Lars can be reached at +49-228-18135361 or lars.hofmann@telekom.de.

Christoph Hartmann holds a M.Sc. in Software Engineering of the Hasso-Plattner-Institute, University of Potsdam. Furthermore, he participated in nine-month design entrepreneurship program at Stanford University. After his studies he joined Deutsche Telekom management program in Bonn, Berlin & San Francisco. He focuses on product innovation and software development. Currently he works at T-Systems International as Technical Project Manager and works on the future IT infrastructure of mobile & fixed net communication. Christoph can be reached at c.hartmann@telekom.de

René Lorenz works as a Management Consultant for Detecon (Schweiz) AG in Zürich. He holds a M.Sc. in Economics and IT and studied International Management in Reutlingen and Mumbai. Prior to join Detecon he worked for the world's largest chemical company and for different consulting firms in Hamburg and Bangalore. René is focusing on customer value management, customer segmentation as well as churn & retention management and is responsible for the customer insight and big data initiatives of Detecon in Switzerland. René can be reached at rene.lorenz@detecon.com

Jan Safka spent, after working in number of positions in IT and business, the last 6 years working for T-Mobile International / Deutsche Telekom Headquarters. Jan is responsible for IT, finance and business technology development within Europe. He has aligned IT landscape and business approach within self-service, opened new channels for customers in European countries and with his international team he is directly responsible for operational performance of self service in selected European countries and for business development in CS areas of technology (CRM, routing, self-service…) where he tries to transfer customer experience, sales and efficiency into technology. Jan Safka has studied Economy in Prague and in Cologne and has Ph.D. in computer science.

9 The Company

We make ICT strategies work

Detecon is a consulting company which unites classic management consulting with a high level of technology expertise.

Our company's history is proof of this: Detecon International is the product of the merger of the management and IT consulting company Diebold, founded in 1954, and the telecommunications consultancy Detecon, founded in 1977. Our services focus on consulting and implementation solutions which are derived from the use of information and communications technology (ICT). All around the globe, clients from virtually all industries profit from our holistic know-how in questions of strategy and organizational design and in the use of state-of-the-art technologies.

Detecon’s know-how bundles the knowledge from the successful conclusion of management and ICT projects in more than 160 countries. We are represented globally by subsidiaries, affiliates, and project offices. Detecon is a subsidiary of T-Systems International, the business customer brand of Deutsche Telekom. In our capacity as consultants, we are able to benefit from the infrastructure of a global player spanning our planet.

Know-how and hands-on expertise

The rapid development of information and telecommunications technologies has an increasingly significant influence on the strategies of companies as well as on the processes within an organization. The subsequent complex adaptations affect business models and corporate structures, not only technological applications.

Our services for ICT management encompass classic strategy and organization consulting as well as the planning and implementation of highly complex, technological ICT architectures and applications. We are independent of manufacturers and obligated solely to our client's success.

Detecon International GmbH
Sternengasse 14 - 16
50676 Cologne
Phone: +49 221 9161 0
E-Mail: info@detecon.com
Internet: www.detecon.com