REDDAL

Big Data – harnessing the big potential

Reddal Insights — 27 August 2015
Per Stenius, Seoweon Yoo

Big Data has been a big deal since the early 2010s. Although the biggest hype has already faded, now is an interesting time to dig up this topic again, as many industries have sobered up to the realities of this originally somewhat overhyped buzzword and the first batch of winners (and losers) of implementing Big Data strategies have emerged.

This article originally appeared as a post in LG CNS BLOG (www.lgcnsblog.com) on 28.04.2016. Reproduced here with the kind permission of LG CNS.

By now most people have heard about Big Data – it is and has been a big deal since the early 2010s, although the biggest hype has already faded. Nevertheless, now is an interesting time to dig up this topic again, as many industries have sobered up to the realities of this originally somewhat overhyped buzzword and the first batch of winners (and losers) of implementing Big Data strategies have emerged. This article seeks to introduce some use cases, and building on those, provide some thoughts on how to approach Big Data going forward.

Understanding the big data hype

The term “Big data” was coined by Roger Magoulas, director or market research at O’Reilly Media in 2005, to refer to datasets that were too big to process with existing business intelligence tools[1]

The reason that Big Data has become a big issue in recent years is because of changes in the nature of data, driven by how digital technology is evolving. The difference between traditional data analytics and big data is commonly illustrated as the 4 V’s; volume (the scale of data is exponentially larger), variety (the range of data types and sources greatly vary),

velocity (the speed of data creation and analysis is much faster), and veracity (the uncertainty of data is higher).



Figure 1 – Illustration of the four V's of Big Data with examples[2]

The underlying sources driving the four V's are several. The mobile revolution, the Internet of Things, and the widespread use of social networks are all trends that have led to this evolution. The mobile revolution brought the Internet into our hands wherever we go via smart phones, tablets, and smart devices. We are constantly consuming and producing data, in unimaginable quantities compared to a decade ago. The technological advancement into Internet of Things means that sensors creating, responding to and transmitting data are increasing in rapidly. Things as mundane as toasters, sneakers, and office building scan feed continuous streams of data. Social networks such as Facebook, Twitter, Instagram all encourage individuals to continuously contribute, update and react on data. Millions of users are constantly uploading their own sentiments on Twitter, pictures of what they had for dinner on Instagram, and videos on Youtube. This explosion in social media expands both the quantity as well as the different forms of data. No surprise then that data traffic is growing at an unprecedented rate – even outpacing the famous Moore's law, as estimates are made that 100 times more data will be processed in 2020 compared to 2010[3].

In order to make sense of this explosion of data, new technologies have emerged to support analysis of or across larger datasets. Hadoop, which became an open source project in 2007, is considered to be the mother of all Big Data platforms. However as many limitations of this platform have been identified, a new Big Data project called Spark is about to become "the next big thing" (technically these are not mutually exclusive – many solutions require installing Spark on top of the Hadoop platform[4]). While the demand for Big Data analysis is growing and applications are becoming commonplace, it is still quite expensive to implement a fully fledged Big Data infrastructure. Because of this, Big Data as a Service (BDaaS) is emerging as the cost-efficient way to reap the benefits without having to pay for the full infrastructure[5].

The Big Data hype is dying down (and it is a good thing)

Big Data was a large part of the technology discussion in 2012, with many talking about its endless potential. However three years later, the initial hype has somewhat died down and businesses are beginning to see Big Data for what it really is. There have been constant voices of concern about the difficulty in measuring and ensuring positive ROI of Big Data strategies[6], yet there is no sign of companies slowing down investing and setting up plans for their Big Data transformation[7]. According to Gartner, Big Data has gone past its "peak of inflated expectations," and has entered the "trough of disillusionment." This in layman's terms means that the Big Data bubble has burst and now is the time for practical business applications of Big Data to emerge. This also implies that plans are getting more concrete and realistic, and with that true impact on business can be achieved.



Figure 2 – Gartner's Hype Cycle 2014[8]

Making use of Big Data - focusing on practical issues, rather than the concept

It is clear that the traditional approach to collecting and analysing data is not fit for the fast-changing digital world. Big Data is needed, but how to use it and what to expect is not always easy to fully fathom.

One key reason why Big Data feels so hard to implement is because of the emphasis on the sheer volume of data and the technology it requires. For sure the amount of data is growing extremely fast, and scalability requirements may be hard to define. Sophisticated technologies are available, but they come with a hefty price tag. In addition, the media is always talking about how finding the right hires to work with Big Data is like recruiting the next superstar sports player – one consulting firm even went as far as estimating that there will be a shortage of about 140,000 to 190,000 Big Data engineers in the US alone in 2018[9]. No wonder companies feel they do not have the resources, or the strength to shake up existing well-working business practices to set up and integrate new and expensive Big Data infrastructure, processes and staff into their business.

Indeed, the focus should not be on Big Data per se, but rather on finding practical and often relatively focused (and thus manageable) solutions to existing problems. Companies should seek to learn to utilize Big Data as a tool, rather than to take it as a new business endeavour. This can be illustrated by reviewing a few cases of companies having successfully implemented Big Data. For the purpose of simplicity, we can define Big Data use cases into two major categories – those seeking cost reduction and those driving revenue growths.

Big Data uses cases for reducing costs - knowing where to focus

PASSUR Aerospace Inc., a company that provides data consolidation, information, decision support, predictive analytics, and collaborative solutions for the aviation industry, was able to eliminate the inefficiency of manually estimating the time of arrival of aircrafts through a Big Data solution which they named RightETA (Estimated Time of Arrival). Before the introduction of RightETA, pilots would estimate the time of arrival and let the land staff know. Usually this estimation would be a few minutes off, which led to inefficiency of either the land staff idly waiting around, or the airplane passengers being trapped in the plane until the land staff arrived. PASSUR utilized historical data of flight landings, its own sensory data of air traffic, and also included publicly available data sources such as the weather information in order to precisely estimate the arrival time, lifting the burden of manual estimation from the already busy pilot.

Tesco has developed an analytics model that utilizes historical buying data, expected temperature trends and weather patterns to properly stock store shelves. This tool reduced out of stock for "good weather products" (like barbeque meat) by a factor of four. The analytics system also tracks the effectiveness of all promotions, giving insights for future marketing plans[10].

Union Pacific Railroad was able to decrease derailments by 75% by using predictive analytics

and tools such as visual sensors and thermometers to detect imminent problems and take care of them before a potential derailment.

A key theme that is common across all of these cases is the replacement of intuition with rigorous data analysis. Each company identified a specific area where inefficiency existed due to human error or a manual/missing process. PASSUR found inefficiency in the pilot's estimate; Tesco saw an error in the store manager's intuition on which items to stock. Before Union Pacific Railroad implemented their Big Data strategy, it is probable that seasoned staff would use a gut feeling of which rails to check at which time of the year, based on their memories of derailments in the past. Identifying the area where a more detailed and rigorous analytical approach can bring concrete benefits is an important first step.

Businesses should also organize and utilize its existing internal data sources, and then look for other available data that can support a more detailed solution approach. More often than not, unorganized and siloed internal company data is the biggest obstacle. For example, 35% of Korean companies claim that combining internal data is their main concern in adopting Big Data projects[11]. Incorporating publicly available datasets or adding new ways of tracking complementary data is another key step in the process.

Big Data use cases for increasing revenue - respecting the customer's privacy concerns

Amazon famously has a unique, customized homepage for each and every user, based on their purchase history, shopping bag contents, and browsing history. This leads to increased revenue per customer, since its recommendations are highly relevant to their customers' interests. However a potential drawback is that the recommendations are so tightly linked to a customer's past purchase, making discovery of a new type of product category within Amazon's homepage quite difficult sometimes.

Target provides customized coupons based on customers' purchase history, and has sophisticated knowledge regarding the stage of life of its customers. The infamous story about Target discovering a 16-year-old teenager's pregnancy even before her father, and sending maternity-related product coupons (which enraged her unknowing father), shows how Big Data can create privacy concerns that can lead to a negative impact on brand image. Nonetheless, this customized coupon strategy was able to significantly increase revenue for Target's "Mom and Baby sales[12]

Walmart enables inventory optimization through providing basic social media analytics to local store managers. By monitoring social media buzz during college football season, Walmart is able to determine when discussion about college football in a certain locality is beginning to heat up. This lets them know when they should be promoting certain products that are related to the season and local teams, and stocking them up.

In these cases Big Data provides a personalized customer experience that seeks to maximize revenue per customer. However, as shown by the Target case, one must exhibit care with privacy issues and the security of customers' personal information. Not having a clear and safe guideline scan gets the company in major trouble both in brand image, as well as with

the law.

To avoid being perceived as a “creepy” company that snoops around personal (potentially embarrassing) data, it is important to focus on customer communication. Companies should be as clear about (a) which data is collected, and (b) how it is used. A good practice is to go beyond the long and often unread “Terms and Conditions,” and actually show a list of data points and their use. Amazon has a “Why recommended?” below every recommended item, which shows exactly what data points were used to generate the recommendation.

Another point to keep in mind is that quality should be valued over quantity of collected customer data. In data analytics, the saying “Garbage in, garbage out” is often used to emphasize that the quality of input data determines the quality of results. Simply dumping in large gigabytes of low-quality data will only render low-quality results. Going back to Amazon’s recommendation function, customers can also “Improve recommendations” by editing the collected data. It’s actively encouraging the customer to provide more high-quality data to provide a better service. The lesson to be learned here is that being as transparent as possible in data collection can also enhance data quality.

Big benefits are out there

Although the initial hype is gone, there is no doubt that Big Data is here to stay and its applications will evolve rapidly going forward. Gartner predicts that it is going to take 5-10 years for Big Data to reach the “plateau of productivity.” Wise businesses will use this time effectively. Some businesses are already showing success, utilizing Big Data to get ahead of competition – having the skills required and some experience from real life efforts is becoming a strategic asset quickly, and is soon to be a must-have.

By remembering to start small with a focused, well chosen, and manageable problem, and making sure to consider potential ethical issues, businesses can take on the seemingly insurmountable mission of implementing Big Data and making it a practical success.

Further reading and references

This blog is based on a broad range of articles and reports. I list some of the more interesting here:

https://datafloq.com/read/3vs-sufficient-describe-big-data/166

http://www.businesswire.com/news/home/20150615005197/en/Global-Study-Reveals-90-Percent-Companies-Benefit#.VX-H5_mqpBc

http://business.inquirer.net/192443/most-companies-drowning-in-information-study-says

http://www.pymnts.com/news/b2b-payments/2015/big-data-becomes-industry-specific/#.VX-WH_mqpBc

http://bits.blogs.nytimes.com/2015/06/15/companies-are-moving-on-from-big-data-technology-hadoop/?_r=0

http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

http://smartdatacollective.com/bernardmarr/321911/big-data-service-bdaas-next-big-thing

http://blogs.gartner.com/svetlana-sicular/big-data-is-falling-into-the-trough-of-disillusionment/

http://www.forbes.com/sites/gilpress/2014/08/18/its-official-the-internet-of-things-takes-over-big-data-as-the-most-hyped-technology/

http://www.washingtonpost.com/local/education/colleges-are-using-big-data-to-identify-when-students-are-likely-to-flame-out/2015/06/14/b2cc68f8-03e4-11e5-bc72-f3e16bf50bb6_story.html

http://www.computerworld.com/article/2935588/big-data/aiming-for-secs-big-data-project-sungard-and-google-bet-on-the-cloud.html

https://fortune.com/2015/06/10/ibm-big-data-tech/

http://www-03.ibm.com/press/us/en/pressrelease/46330.wss

http://www.informationweek.com/software/enterprise-applications/big-data-6-real-life-business-cases/d/d-id/1320590

http://www.businesscloudnews.com/2015/05/20/data-as-a-service-provider-delphix-buys-data-masking-specialist-axis/

http://venturebeat.com/2015/06/03/data-science-cloud-service-plotly-raises-5-5m/

http://www.datanami.com/2015/06/12/8-new-big-data-projects-to-watch/

http://www.forbes.com/sites/danwoods/2011/11/03/explaining-hadoop-to-your-ceo/

http://www.breachlevelindex.com/#!breach-database

http://www.wsj.com/articles/SB10001424052702303722104579240140554518458

http://wi.mit.edu/news/archive/2013/scientists-expose-new-vulnerabilities-security-personal-genetic-information

http://www.wired.com/2010/03/netflix-cancels-contest/

[1]http://www.quora.com/Who-coined-the-term-big-data

[2]http://www.ibmbigdatahub.com/infographic/four-vs-big-data

[3]http://www.computing.co.uk/ctg/news/2334702/moore-s-law-can-t-help-us-now-big-data-processing-and-the-drain-on-energy-supplies

[4]http://www.forbes.com/sites/bernardmarr/2015/06/22/spark-or-hadoop-which-is-the-best-bi

g-data-framework/

[5]http://smartdatacollective.com/bernardmarr/321911/big-data-service-bdaas-next-big-thing

[6]http://www.informationweek.com/big-data/big-data-analytics/big-data-roi-still-tough-to-measure/d/d-id/1110150

[7]http://www.businesswire.com/news/home/20150615005197/en/Global-Study-Reveals-90-Percent-Companies-Benefit#.VX-H5_mqpBc

[8]http://www.gartner.com/newsroom/id/2819918

[9] http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

[10]http://www.rfgen.com/blog/bid/285148/Tesco-Improves-Supply-Chain-with-Big-Data-Automated-Data-Collection

[11] Micro Strategy Korea & Korea IDG, 2014.9

[12]http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=1&_r=2&hp