BIG DATA: BIG IS THE NEW SMALL

Gayan Lakmal Alwis

Senior Portfolio Manager
Ceylinco Life Insurance
Limited

1. What is the big deal about data?

Data lies before us, behind us, around us, between us, and eventually, within us.

Over the last decade, the amount of data generated has been exploding, at an exponential rate. Companies capture trillions of bytes of information about their customers, suppliers, and operations, and millions of networked sensors are being embedded in the physical "smart" world in devices such as mobile phones and automobiles – sensing, creating, communicating and archiving data. Multimedia and individuals with smartphones and on social network sites will continue to fuel this exponential growth.

The data flow so fast that the total accumulation of the past two years – which amounts to a zettabyte[1] – dwarfs the prior record of human civilization. It is said[2] that there is a "big data revolution" – but it is not the quantity of data that is revolutionary. The revolution is that now we can do something with those data.

To put this in to context, a report by the International Data Corporation (IDC)[3] predicts that, from 2005 to 2020, the global data volume will grow by a factor of 300, from 130 exabytes to 40,000 exabytes, representing an annual growth rate of over 46 percent. For ease of comparison, we could note that if all words spoken by humans were digitized as text, they would total only about 5 exabytes![4]. In a different setting, a McKinsey Report[5] states that the potential value of global personal location data is

[1] A zettabyte is equal to 1,180,591,620,717,411,303,424 bytes.

[2] Professor Gary King. See http://harvardmagazine.com/2014/03/why-big-data-is-a-big-deal

[3] J. Gantz and D. Reinsel, "The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east,'' in Proc. IDC iView, IDC Anal. Future, 2012

[4] Peter Lyman and Hal R. Varian, How much information? 2003, School of Information Management and Systems, University of California at Berkeley, 2003

[5] J. Manyikaet al., Big data: The Next Frontier for Innovation, Competition, and Productivity. San Francisco, CA, USA: McKinsey Global Institute, 2011, pp. 1-137.

estimated to be USD 100 billion in revenue to service providers over the next ten years and be as much as USD 700 billion in value to consumer and business end users.

Interestingly, total global storage and computing capacity from 1986 to 2007 grew at an annual rate of 23 percent over that period[6]. Highlighting the rise of digitization, the percentage share of data stored in digital form increased from only 25 percent in 2000 to a dominant 94 percent in 2007.

Figure 1: Evolution of Storage Space

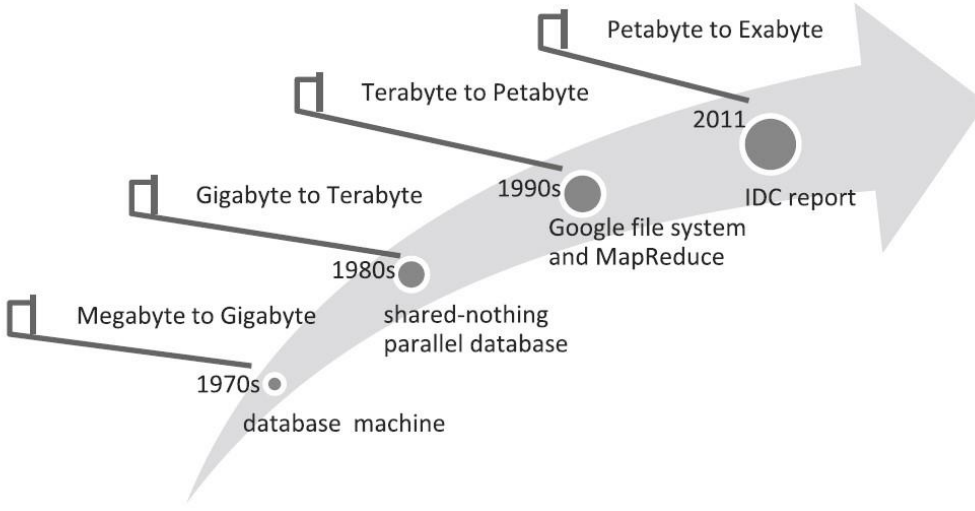


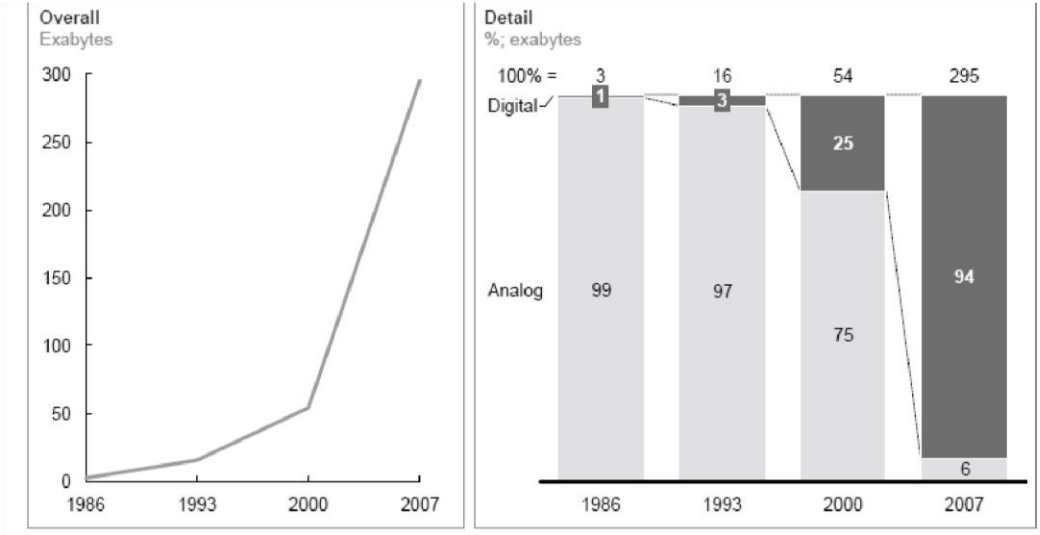
Table 1: Storage Metrics

Unit	Value	Value in bytes
kilobyte (KB)	2^{10} bytes	1,024
megabyte (MB)	2^{10} x 2^{10} bytes	1,048,576
gigabyte (GB)	2^{10} x2^{10}x 2^{10} bytes	1,073,741,824
terabyte (TB)	2^{10} x 2^{10} x 2^{10}x 2^{10} bytes	1,099,511,627,776

[6] Martin Hilbert and Priscila López, "The world's technological capacity to store, communicate, and compute information," Science, February 10, 2011.

petabyte (PB)	2^{10} x2^{10}x 2^{10} x 2^{10} x 2^{10} bytes	1,125,899,906,842,624
exabyte (EB)	2^{10} x 2^{10} x 2^{10} x 2^{10}x2^{10} x 2^{10} bytes	1,152,921,504,606,846,976
zettabyte (ZB)	2^{10} x 2^{10} x 2^{10} x2^{10} x 2^{10} x2^{10} x 2^{10} bytes	1,180,591,620,717,411,303,424

Table 2: Globally installed (optimally compressed) storage



NOTE: Numbers may not sum due to rounding.
SOURCE: Hilbert and López, "The world's technological capacity to store, communicate, and compute information," *Science*, 2011

2. What is it about data that is big?

Big Data is one of those buzzwords that confounds and puzzles as to what it means in the real world. A careful study into its origins and literature brings about three interesting definitions.

Attributive definition

"Big data technologies describe a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and/or analysis''[7].

[7] J. Gantz and D. Reinsel, ``Extracting value from chaos,'' in *Proc. IDC iView*, 2011, pp. 1-12

This definition delineates three salient features of big data, which has become popular cornerstones in the Big Data literature and architecture – volume, variety and velocity. As a result, the "3Vs"' definition has been used widely to characterize big data.

- Volume: big data does not sample, most of the time it just observes the entire population and track what happens. The size of the data determines the value, potential insight and most of the time, whether it can actually be considered big data or not.

- Variety: unlike traditional data which are typically structured and can thus be easily tagged and stored, the vast majority of today's data, such as user-generated content (text, images, audio and video) are unstructured. This is further elaborated in section 4.
- Velocity: the speed at which data is generated and processed to meet the demands and challenges that lie in the path of growth and development. Big data is often available in real-time.

Additionally, a new V for "Veracity" is added by some organizations to describe the biases, noise and abnormality in data and to highlight that the quality of captured data can vary greatly, affecting accurate analysis.

Table 2: Big Data Differentiators[8]

Dimension	Challenges of Traditional Data Management Techniques	Big Data Differentiators
Volume	• Traditional analytics are often designed to analyze relatively small sample sizes. • Data storage across multiple drives presents problems for traditional techniques. • The cost to analyze large data sets using traditional techniques is too high, both in time and memory.	• Big Data techniques are designed to handle huge amounts of data spread across multiple storage devices and platforms. • Big Data technologies facilitate massive parallel processing for faster access and analytics. • Low-cost storage and cloud storage are increasingly available.

[8] "Where have you been all my life? How the financial services industry can unlock the value in Big Data," PwC FS Viewpoint, October 2013.

Variety	• The proliferation of data types and models creates compatibility issues with traditional tools. • The increasing demand for data mashups and deep insights challenges traditional data techniques that struggle with nonnumerical data.	• Big Data frameworks are designed to accommodate varying data platforms and data models. • Advanced technology stacks are designed to provide insightful analysis on a diverse range of structured and unstructured data sets.
Velocity	• Rapidly updating data sets require dynamic, real-time analysis that is not available with traditional • techniques. Information management processes need to intelligently decide in real time what data to save and what to discard.	• Big Data techniques that dynamically analyze data in near real-time can efficiently update results based on new information. • Advanced algorithms can identify useful data to keep versus lowvalue data to discard so as to appropriately address storage needs.

Comparative definition

"Datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze"[9].

This definition is subjective and does not define big data in terms of any particular metric.

However, it incorporates an evolutionary aspect in the definition (over time or across sectors) of what a dataset must be to be considered as big data.

Architectural definition

"Big data is where the data volume, acquisition velocity, or data representation limits the ability to perform effective analysis using traditional relational approaches or requires the use of significant horizontal scaling for efficient processing"[10].

However, reaching a consensus about the definition of big data is difficult, if not impossible and coincidentally – not required for our purpose as well. A logical choice

[9] J. Manyikaet al., Big data: The Next Frontier for Innovation, Competition, and Productivity. San Francisco, CA, USA: McKinsey Global Institute, 2011, pp. 1-137

[10] M. Cooper and P. Mell. (2012). *Tackling Big Data*. See: http://csrc.nist.gov/groups/SMA/forum/documents/june2012presentations/fcsm_june2012_cooper_mell.pdf

might be to embrace all the alternative definitions, each of which focuses on a specific aspect of big data.

3. How big means "Big"?

To know how "big" the big data can become, a brief look at few social media statistics[11] would suffice.

Internet	As of July 2015, total world population was 7.3 billion. The internet had 3.17 billion users, whereas there were 2.3 billion active social media users. Internet users had an average of 5.54 social media accounts.
Google	Google processes 100 billion searches a month (which is equivalent to an average of 38,000 search queries every second) and those searches are carried out by 1.17 billion unique users. Google has answered 450 billion unique queries since 2003 and more than half of Google's searches come from mobile devices. By 2014, Google has indexed 30,000,000,000,000 pages of the internet.
Facebook	Facebook adds 500,000 new users every day; 6 new profiles every second. 72 percent of all online US adults visit Facebook at least once a month. About 4.5 billion likes are generated daily, while photo uploads total 300 million per day[1]. The average number of friends is 338, and the median number of friends is 200, while the average time spent per Facebook visit is 20 minutes. For every 60 seconds, 510 comments are posted, 293,000 statuses are updated, and 136,000 photos are uploaded.
Twitter	Twitter has 1.3 billion accounts, but only 320 million are active. About 500 million people visit Twitter each month without logging in. The average Twitter user has 208 followers but 391 million accounts have no followers at all. There are 500 million Tweets sent each day. That's 6,000 Tweets every second. 65.8% of US companies with 100+ employees use Twitter for marketing and 77% of Twitter users feel more positive about a brand when their Tweet has been replied to.
YouTube	300 hours of video are uploaded to YouTube every minute and there are 3.25 billion hours of video watched each month while more than half of views come from mobile devices. The average mobile viewing session lasts more than 40 minutes.

Now add to this other relatively less popular *(in Sri Lankan context)* social media networks such as LinkedIn, Google+, Weibo, Flickr, Myspace, …. and the list goes on!

[11] See https://www.brandwatch.com/2016/03/96-amazing-social-media-statistics-and-facts-for-2016/

4. The tale of two terms: Structured vs. Unstructured Data

Structured Data

The term "structured data" generally refers to data that has a defined length and format. For example, customer's name, address, or contact number. Structured data are usually stored in a database and we can query it using a language like Structured Query Language (SQL). All firms collect structured data from "traditional" sources and might include Customer Relationship Management (CRM) data, Enterprise Resource Planning (ERP) data, and financial data.

Although this might seem like business as usual, in reality, structured data is taking on a new role in the world of big data. The evolution of technology provides newer sources of structured data being produced – often in real time and in large volumes. The sources of data are divided into two categories:

Machine-generated: refers to data that is created by a machine without human intervention. Machine-generated structured data can include the following:

I. Sensor data: examples include Radio Frequency ID (RFID[12]) tags, smart meters, medical devices, and Global Positioning System (GPS) data.
II. Web log data: when servers, applications and networks operate, they capture all kinds of data about their activity. This can amount to huge volumes of data that can be useful, for example, to deal with service-level agreements or to predict security breaches.
III. Point-of-sale data: when the cashier swipes the bar code of any product that we are purchasing, all that data associated with the product is generated.
IV. Financial data: many financial systems are now programmatic. They are operated based on predefined rules that automate processes. Trading data of a stock is a good example.

Human-generated: data that humans generate, by interacting with computers. Examples of structured human-generated data might include the following:

[12] Radio-frequency identification (RFID) uses electromagnetic fields to automatically identify and track tags attached to objects. RFID is rapidly becoming a popular technology. It uses tiny computer chips to track items at a distance. An example of this is tracking containers of produce from one location to another.

I. Input data: this is any piece of data that a human might input into a computer, such as name, age, income, survey responses, etc. This data can be useful to understand basic customer behavior.
II. Click-stream data: data is generated every time we click a link on a website. This data can be analyzed to determine customer behavior and buying patterns.

III. Gaming-related data: every move we make in a game can be recorded.This can be useful in understanding how end users move through different levels in a game.

Unstructured data

Unstructured data is data that does not follow a specified format. It is estimated that 80 percent of the data available to enterprises amounts to unstructured data. Unstructured data is really most of the data that we would encounter in our day to day life. Until recently, however, the technology did not exist to extract value out of it, except storing it or analyzing it manually.

Unstructured data is everywhere. In fact, most individuals and organizations conduct their lives around unstructured data. Just as with structured data, unstructured data is either machine generated or human generated.

Machine-generated unstructured data includes:

I. Satellite images: includes weather data or the data that the government captures in its satellite surveillance imagery.
II. Scientific data: includes seismic imagery, atmospheric data, and high energy physics.
III. Photographs and video: includes security, surveillance, and traffic video.
IV. Radar or sonar data: includes vehicular, meteorological, and oceanographic seismic profiles.

Human-generated unstructured data includes:

I. Enterprise information: all the text within documents, logs, survey results, and e-mails. This actually represents a large chunk of the text information in the world today.
II. Social media data: generated from the social media platforms such as YouTube, Facebook, Twitter, LinkedIn, Flickr, etc.

III. Mobile data: includes data such as text messages and location information.IV. Website content: comes from any site delivering unstructured content.

5. Size really does matter

Big data invariably contains large amounts of data, but how much is large? There are no straight answers, but if the amount of data is breaking traditional systems, then it is Big Data. Few key Big Data principles to keep in mind are:

- Big Data solutions are ideal when all, or most, of the data, i.e. the entire population, needs to be analyzed versus a sample of the data; or a sampling of data is not nearly as effective as a larger set of data from which to derive analysis.

- Big Data solutions are ideal for analyzing not only raw structured data, but semi-structured and unstructured data, especially from a wide variety of sources.

- Big Data solutions are ideal for iterative and exploratory analysis when business measures on data are not predetermined.

Along with this "big" expansion of the availability of "data", the natural questions arises – what does this means to us? Can it change the game? And more importantly – what are we going to do about it?

6. The big picture: What's in it for me?

Big data analysis played a large role in Barack Obama's successful 2012 re-election campaign[13] and was in part responsible for the BJP to win the Indian General Election - 2014[14]. While there are well documented and researched applications/uses of Big Data covering a wide spectrum of fields of interest, this section especially outlines several real-life, Big Data use cases, which generate tremendous value to respective financial institution.

[13] Lampitt, Andrew. "The real story of how big data analytics helped Obama win". See http://www.infoworld.
com/article/2613587/big-data/the-real-story-of-how-big-data-analytics-helped-obama-win.html 14 "News: Live Mint". Are Indian companies making enough sense of Big Data?
See http://www.livemint.com/Industry/bUQo8xQ3gStSAy5II9lxoK/Are-Indian-companies-making-enoughsense-of-Big-Data.html

Segment 1: Sentiment analytics

Social media has opened new avenues and opportunities for organizations to connect with their customers, but the sheer volume of communications about brands, products and services; discussed, shared, criticized or liked on different social platforms can be overwhelming.

Sentiment analytics helps to rapidly read all this data, providing an executive summary of what people like and do not like about a company brand or products. The reasons behind the sentiment can then be easily extracted, providing valuable business insights.

Objective	Business Intellect	Role of Big Data	Use Case
Monitor what customers say to adjust marketing tactics	Meaningful data includes opinions, feelings and attitudes about a brand, topic or keyword, which are shared freely in the world wide web. K nowing the opinion and attitude of the customers offers profound knowledge in order to adjust marketing tactics correctly.	An easy source of customer sentiment is from the social sphere, including social networks, blogs and review sites. This data is naturally unstructured and dynamic as new text is generated continuously. This data is then suited to measuring sentiment over time such as before and after an organization's branding efforts. Internal data gathered from past consumer surveys and call logs may also provide a good source to measure the customer sentiment towards particular products.	Nedbank Ltd is a large bank in South Africa that realizes great advantages by using social media analytics[2]. Analyzing various social media platforms in almost real-time provides Nedbank's marketing department information about the marketing campaign, customer preferences and complaints. This technology implementation has decreased social media monitoring costs significantly while enhancing marketing success.

Identify key customers to boost word-ofmouth Marketing	The right sentiment analysis tool can also identify the most influential customers regarding company brands or products. It enables to engage with the right people who will spread the word and has an influence on a social platform. Those key customers are critical in order to fulfil the goals for a successful acquisition strategy.	Internal customer profile may reveal which customers or clients give a high number of referrals. This database may also contain the attributes that can be used to determine if the customer is influential, such as demographic information. Link analysis strongly depends on highly unstructured social network data and data from third party blogs and review sites.	T-Mobile understands that customers can sometimes fall like dominoes[3]. It only takes one highly influential customer to switch to another network provider before large numbers of customers follow suit. To try and find out who the valuable high influential customers are, T-Mobile uses a mixture of customer influence scoring and customer value KPI's to build an influence profile of each customer[4]. They can then target these customers for special attention and take appropriate actions.
Examine customer feedback To improve products and services	Many consumers freely give feedback and product suggestions on social media websites. Big Data technologies can be used to identify those valuable consumer insights to improve products and services much faster than with traditional surveys, which only portray a small sample group at one specific moment in time.	The social sphere including the own Facebook homepage and twitter can form the basis of data source for customer sentiment. Twitter is particularly useful as its openness allows searching all tweets across the entire network rather than those just on the homepage.	Barclays was able to derive actionable insights from realtime social media analytics after they launched their new mobile banking app. The app did not allow young consumers under 18 years to transfer or receive money. This created negative comments from teenagers as well as from their parents as they could not transfer any money to their kids. After the data revealed this problem, Barclays improved their app promptly, adding access for 16 and 17-yearold's[5].

Segment 2: Customer 360

Understanding the customer as a whole is important to stay ahead of competitors.

There are several important aspects to consider when developing a 360-degree customerview. The past and immediate customer behavior is important to predict future customer trends and what their most likely next action will be. The customer's transactions, including eating and travel habits are also important to build a lifestyle profile and discover new insights. These are just some customer attributes that are used to build a complete and holistic customer picture.

Objective	Business Intellect	Role of Big Data	Use Case
Identify the customer Profile	Knowing the customer profiles means having a deep understanding about them, which can be used to drive actionable insights. This can lead to improved marketing campaigns, targeted sales and better customer service. A clearer view about the customer profile enables companies, for example, to send out triggered messages, which is a good way to reinforce the brand and target customers.	Customer profile is primarily used to add personalization to marketing messages. This includes the customers' demographics such as their post code and their name. Transactional data may also reveal the customers' interests, which can be used to help add relevant information to the marketing messaging.	HDFC bank is the 5th largest bank in India by assets. HDFC Bank uses customer lifecycle events to boost credit card activations[6]. This is achieved by targeting promotions with personalized messages to each of the lifecycle segments that HDFC had identified. The result of this is a significant increase in the number of credit card activations and a reduction in cost per acquisition of each customer.
Understanding the product engagement of the customer	Understanding how consumers are using a specific product and then making decisions accordingly makes a big difference. With the help of Big Data analytics companies can find out how engaged a consumer is with a product. This can help to send the correct marketing message and product when the customer needs or wants it most.	To accurately assign a product lifecycle stage to a customer, profile data and transaction data is needed to determine how they use the products. Metrics such as customer engagement can be gained from this data. Recent clickstream URL referrals and cookie data gives information if the customer is in the research stage of the product cycle.	The First Tennessee Bank uses product lifecycle stages such as loyal customers, high value customers, apprehensive customers and wavering customers to name just a few key segments[7]. For example, reengagement messaging may be useful for past high value customers who have not been in contact with the bank for a while loyal and engaged customers may prefer more VIP offers.

Detect when a customer is about to leave	Analyzing customer behavior is not just reviewing past historical purchases, but a tool to forecast future actions and trends of customers. By predicting customer behavior, insights can be revealed to stop churn before it is too late.	Customer profile and transaction data can be used to analyze how the customer is using the product, which in turn can be used as behavioral indicators of potential churn. Social data from sentiment analysis may also indicate if particular segments are unhappy with an organizations service and are likely to churn.	Customer data of T-Mobile USA includes the time and lengths of call, internet usage or peak times for direct messaging. T-Mobile takes advantage of this data to prevent customer churn[8]. An example of this is billing analysis, where the product usage is calculated. If the frequency of calls to contacts who are using a new provider are increasing this could imply that friends or family are switching providers, and the customer might possibly do so as well. By identifying these customers T- Mobile achieved a decrease their churn rate by 50% in just one quarter.

Segment 3: Customer segmentation

Segmentation is simply dividing the customers into natural groupings that share similar characteristics or behaviors.

Understanding these groups is necessary to find needs and wants that form the basis for a sales and marketing strategy. The concept of segmentation is not new and, frequently used in companies, but Big Data facilitates to create sharper segments faster.

It helps to see the existing customer base in new ways, which creates unique business opportunities. It has made possible, what is simply called “segment of one” marketing, where targeted, unique product offerings are made individually, to each targeted customer.

Objective	Business Intellect	Role of Big Data	Use Case

Design targeted marketing programs	Big Data segmentation allows to see how customers are really using products and what issues they care about most. This enables companies to discover segments that have traditionally been underserved. Highly optimized marketing messages for each of the groups can then be developed, creating greater resonance with customers.	All customer data can be used for cluster analysis, but better results would be achieved if the data is focused to a specific demographic or product users. An example would be to investigate students that use savings accounts. For that the retiree demographic would not add any meaningful insights.	By segmentation, Bank of America is able to remove its assumptions about its customers[9]. This led to a change in its marketing message from *"use the value of your home to send your kids to school"* to *"use the value of your home to do what you always wanted to do"* and increased conversion rates tenfold.
Creating loyalty programs based on card usage habits	Accurately defining segments enables banks to provide highly personalized cash back offers with vertical partners such as food stores, retail or travel companies. This can significantly increase card loyalty and card usage.	Transaction data is primarily used for this use case, but customer profile data can help narrow the aim is to design a loyalty program for a particular demographic.	Citibank Singapore offers customer discounts at retailers and restaurants based on the customer transactional patterns[10]. By offering this service, Citibank has a significant increase in its card usage loyalty, retention and overall improvement of customer satisfaction.
Optimize pricing strategy	Knowing how much a segment is willing to pay for a product or service is a key business strategy for many companies. Big Data segmentation allows to find more of these groups and their price willingness, creating a sliding scale optimized pricing strategy for the customer base.	Transaction history and customer's profile is needed to view the customer's behavior and what their spending potential is. Detailed customer profile can allow price clustering based on demographics and lifestyle behaviors.	Fifth Third Bank uses analytics-based product pricing engine to help acquire new customers[11]. Using data analytics the Bank can run scenarios on how various price points will influence its customer acquisition and deposit levels. For example, the bank can make price predictions when interest rates will rise in the future and make scenarios where it wants to be with rates in the market to be aggressive in attracting customers.

Build relationships with valuable customers	With the help of Big Data, profitable market groups can be identified and given preferential treatment to strengthen customer satisfaction. Segmentation also reveals the attributes of a profitable person, enabling banks to target high profit potential customers.	Some key data is needed to measure customer profitability. This includes demographics, the type, frequency and occasions the product purchases and the time and type of marketing contacts. This would typically come from marketing and sales data.	Barclays is making use of segmentation by targeting students with personalized advertisement and attractive offers[12]. The objective of reaching out to this market segment is straightforward. Students most likely pick a bank for the first time and Barclays tries to keep these customers in the long term when they will become profitable customers.

Segment 4: Next best offer

Next best offer allows an organization to increase its up-sell and cross-sell opportunities by predicting what the customer wants next.

This is achieved by analyzing the customer's market basket and finding the patterns between products to forecast future purchases. Leveraging this knowledge can ultimately lead to improved marketing return on investment, customer loyalty and sales results.

Objective	Business Intellect	Role of Big Data	Use Case
Enhancing loyalty Through next best offer	Offering a customer an additional product that suits his or her wants and needs will not only increase sales, it can also strengthen customer relationship. Relevant offers might increase the interest and "customer stickiness" to banking products.	Transaction history of the customer is the most important data for this use case. Profile data is also required to help predict the customers' life stage events to optimize the timing of marketing messages.	Netflix is a completely data driven company and can accurately predict what movies its customers like and hate. To achieve this, Netflix feeds all its customer data into a recommendation system that takes into consideration all aspects about the customer, from viewing behaviors to customer profiling. Netflix then use predictions generated from the recommendation engine to make tailored movie recommendations based on each customer's preferences.

Measuring product Propensity	Product propensity scoring can be used to boost revenue by offering customers products or services that they really want.	The customer's propensity towards a certain product can be measured through both their own historical purchases and their peers. This requires customer profile data and their transaction history.	A pioneer in Next best offer is Amazon. The company employs collaborative filtering to predict a customer's product propensity based on the customer's peers. From this, Amazon generates its "you might also want" and "customers also bought" campaigns for each product purchased and viewed by the customer. This has led to a significant increase in Amazon's cross sales and revenue.
Product bundling to uplift revenue	Determining which products are most likely to be purchased together is essential in order to develop highly effective product promotions and to increase revenue.	Customer profile and transaction data is needed to build an effective cross selling model. Care should be taken as many products in financial services are bundled with other products or have been bundled in the past.	Manulife Financial makes use of data analytics to bundle the deposit and credit account[13]. As Canada's first flexible mortgage account, the bank offers a mortgage with checking and savings, using the net balance to calculate the specific interest, which positions the account to serve as the client's primary current account.

Segment 5: Channel journey

With so many ways a customer can interact with a company, including mobile, social media, clickable ads, stores, TV and publication platforms, keeping track of the customer journey can be a difficult task.

Big Data helps by taking a holistic view of the entire customer journey and experiences on each channel. This can be used to find patterns of usage that lead to sales, or where and which channels are underperforming. From this knowledge, banks can optimize the funnel conversion[14], increase messaging effectiveness and measure marketing results across all channels.

Objective	Business Intellect	Role of Big Data	Use Case

[14] Conversion funnel is a phrase used in e-commerce to describe the journey a consumer takes through an Internet advertising or search system, navigating an e-commerce website and finally converting to a sale. The metaphor of a funnel is used to describe the decrease in numbers that occurs at each step of the process.

Provide more relevant content in the preferred channel	Customers may not use all channels in the same way and so messaging on each channel should reflect this. Some segments may prefer to research products on a blog before visiting the website and landing pages could adapt to this by displaying relevant offers and further reading linked to their research.	Data from online usage from cookies, URL referrals and metrics can be used to determine which online channels the customers are using and what they are using them for.	HDFC uses their knowledge of the channel journey to personalize the customers experience on their channel of choice[14]. For example, an ATM would recognize the customer's preferred language from previous interaction on the organization's website. Coupled with other preferences measured across a wide range of channels has decreased time at ATMs by 40% and helped reduce operational costs.

Recognize multi-channel behaviors that lead to sales	Big data analytics allows companies to recognize patterns of channel usage in their customer segments and how successful journeys lead to sales and where unsuccessful channel journeys fail. This then allows to optimize the funnel conversion by optimizing bottlenecks and fixing points of high drop off rates in the customer channel journey.	Customer profile data will be required to view which and how customer segments use channels. Offline metrics such as channel performance, time stamps and any comments made about the customers' visit can be used to build a picture of their offline journey, while online metrics gained from cookies, webpage JavaScript, tracking codes and Matchback are used to build the online journey.	Vodafone recognized their traditional approach of measuring the customer channel journey using cookies was adversely affecting their "return to basket" email campaign[15]. This was due to 40% of their customer base using their phones for emails, which use different cookies than their laptops, so information about their abandoned shopping basket was lost between channels. To solve this, Vodafone deployed deep linking data analytics to track the customer across channels and platforms and they now deliver a completely cookiefree experience. Return to abandoned baskets has since increased by 30%.

Guide customers to low cost channels	Sometimes different channels can provide the same service, but at widely different costs to a company. Aided by Big Data analytics, guiding customers to low cost channels requires robust self- service and a customer driven approach to provide a quality and intuitive service.	Data from online JavaScript and time stamp scan indicate the kind of content and how long the user has been viewing the page. This is then combined with the profile data to determine how each customer is using the channel.	HSBC found the primary barriers for internet banking were customer habit, security concerns and a lack of confidence[16]. They now have an active migration strategy to address these concerns. Part of the HSBC migration strategy is to enable customers to undertake increasingly more complicated banking activities via the internet.
Measure marketing effectiveness across channels	The customer may purchase a product in one channel, but had made the decision to purchase it because they had seen an advertisement in another channel the week before. The question is how much credit should be given to each channel for selling the product?	Measuring online click streams, tracking codes and cookie data is required to build a complete picture of the user's online journey. This is then combined with the customer profile and Ad statistics from all the mediums used to build apicture of channel performance.	Laurentian Bank of Canada uses data analytics to aid its understanding of the performance, factors and influence of its marketing campaigns across its various channels[17]. This allows Laurentian Bank marketing managers to appropriately change its campaigns accordingly when they are not performing well.

7. The Big (Data) Bang: Imagine…. the possibilities

Imagine …. a credit card, which offers amazing discounts to buy what you always dreamt, or offers an easy payment scheme with your favorite retailer or travel agent, to tick off that bucket list.

Imagine …. a paperless credit appraisal system, in which you can apply for an instant loan - from your smart phone, that allows immediately to withdraw the auto-approved loan proceeds, all in few minutes, while you take a walk to the nearby ATM.

Imagine …. a current account, which knows what matters to you most, and honour those cheques, even if you are above your pre-approved overdraft limit.

Imagine …. a current account, which learns about the cheques you frequently draw to meet the utility bill payments, and auto generates an auto subscription to online bill settlement and ask your approval to continue.

Imagine …. a fixed deposit, which auto renews itself, at the best rate possible, until your imported sports car reaches the port.

Imagine …. a foreign currency account, which converts part of your US Dollars to Sterling Pounds, to meet that annual associate member subscription.

Imagine …. a personal loan, which automatically extends you a complimentary grace period, to meet your larger than expected credit card bill, because of the diamond ring you suddenly purchased, to ask that larger than life question.

Imagine …. a Bank that sends you a complimentary travellers insurance and a fine quote to purchase the foreign currency, for the vacation you just booked online.

Imagine …. a Bank which knows exactly how much cash float is needed at branches, for any given day, for any given month.

Imagine …. a Bank that offers a competitive fixed deposit, which is auto-priced in real time, with a "best price (rate) guarantee".

Imagine …. a Bank that reminds you about that "deal of the day", when you walk (or drive) pass your favorite retailer or restaurant.

Imagine …. a Bank that knows when to call you and when to not. More importantly, a Bank that calls you – not the other way around.

Imagine …. a Bank that doesn't have a hotline, call centre or even a receptionist, because they don't need one!

Imagine …. a Bank that offers exactly what you want, at the exact time you want, in the exact way you want it, and keep doing exactly that.

8. Big Data: Is it for me?

To determine if Big Data is the right answer for your firm, PwC[15] sets out the following three guidelines.

[15] "Where have you been all my life? How the financial services industry can unlock the value in Big Data," PwC FS Viewpoint, October 2013.

A. Does your organization have business issues or opportunities that point to Big Data?

i. Have competitors recently adopted Big Data to leapfrog ahead? ii. Have business unit leaders initiated projects where they need new types of business intelligence analysis?
iii. Is the organization able to monitor its image on social media? iv. Would specific lines of business benefit from Big Data-enabled analysis?

B. Is Big Data the right way to solve your problems?

i. Does your institution have the right resources to quickly and easily analyze its vast stores of data?
ii. Do the challenges your business faces lend themselves to proven applications of Big Data, such as pattern recognition, predictive modelling, time series analysis, or visualization?

C. Is Big Data a good fit for your organization?

i. Does your organization have a culture and appetite for risk and innovation? ii. Does your institution have the right resources to quickly and easily analyze its vast stores of data?
iii. Do some of your departments leverage analytics extensively? iv. Does your institution have the right skill set for experimenting with new technologies?
v. Does your organization have experience in evaluating new technologies from technical and operational perspectives in addition to measuring their value?

9. From where to begin?

A. It is important to begin with a real world view of Big Data – what it is and what it isn't Most organizations still struggle to integrate the data they have already collected over the years. Realistically, Big Data alone will not solve that problem, but enabling Big Data will reveal opportunities and lead to potential solutions.

B. Before investing in Big Data, first consider whether your organization is already using its existing data effectively.

Does your organization currently:

- Collect the data you need?
- Analyze what you need?
- Discard what you do not need?
- Distribute what adds value?

If the answers to the above questions are *yes*, then all systems are go for launch. In that case, PwC's framework [16] for response provides the structure of a well-defined approach to Big Data.

1. **Determine if Big Data is the right answer for you**
   Do not explore emerging technology for the sake of academic interest. Assess what the business problem is, whether it is valuable, and if it can be enabled by Big Data.

2. **Design and establish an organization**
   It is imperative that financial institutions take a systematic approach to build a centralized data organization to foster data innovation and agility.
3. **Establish a business case evaluation**
   Create a framework to consistently evaluate benefits, risks, and strategic alignment of Big Data implementations.

4. **Pilot, assess, and operationalize**
   Run research and development experiments on Big Data opportunities, and operationalize if they add value.

5. **Evaluate and improve**
   Determine if the opportunity adds the value that was expected, and look for ways to improve the organization's Big Data capability.

10. Big Data is a big deal!

Data have become an important factor of production today – on par with physical assets and human capital – and the increasing intensity with which enterprises are gathering information alongside the rise of multimedia, social media, and the Internet of Things (IoT) will continue to fuel exponential growth in data for the foreseeable future.

[16] "Where have you been all my life? How the financial services industry can unlock the value in Big Data," PwC FS Viewpoint, October 2013.

Big Data is an inflection point when it comes to information technologies: in short, Big Data *is* a Big Deal! In fact, Big Data is going to change the way you do things in the future, how you gain insight, and how you make decisions (this change is not going to be a replacement for the way things are done today, but rather a highly valued and much anticipated extension) As a word of caution, we should understand that big data may not necessarily lead to big insights. The global financial crisis clearly highlighted what can happen when rich data and analytics collide with gaps or lapses in judgment. Lacking this good judgment, big data can lead to bad decisions and (literally) big disasters.

To overcome this insight deficit, "big data", no matter how comprehensive or well analyzed, must be complemented by "big judgment"[17]. Hence, we should always note that no matter how impressive or aggressive the data and analytics become, we should always rely on human judgment, intellect and values to create a better world.

"The data themselves, unless actionable, aren't relevant or interesting. What is interesting is what we can now do with them to make people's lives better"[18].

References

- H. Hu, et al, (2014), Toward Scalable Systems for Big Data Analytics, IEEE
- Evry, Big Data in Banking for Marketers: How to Derive Value from Big Data
- Minelli, Michael, et al (2013), Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses, Wiley
- Zikopoulos, Paul, et al, (2012) Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, McGraw-Hill
- J. Manyikaet al., (2011), Big data: The Next Frontier for Innovation, Competition, and Productivity, McKinsey Global Institute
- PwC FS Viewpoint, (2013), Where have you been all my life? How the financial services industry can unlock the value in Big Data
- Monahan, Tom, (2013), "Big Data Calls for Big Judgment – in Finance and Beyond" (available at https://www.cebglobal com/blogs/big-data-calls-for-big-judgment-in-finance-and-beyond/)

[17] See https://www.cebglobal.com/blogs/big-data-calls-for-big-judgment-in-finance-and-beyond/

[18] Nathan Eagle, an adjunct assistant professor at Harvard School of Public Health. See http://harvardmagazine.
com/2014/03/why-big-data-is-a-big-deal