

Big Data Phenomenon in Banking

Article by Bernard Aliu

Department of Information Technology, Texila American University

E-mail: albernicus@yahoo.com

Abstract

The banking industry has significantly thrived with the rapid advancement of human civilization and the associated growth in information technology. Earlier to this digital era, all the transactions and the business intelligence thereof had excessive human involvement. The digital world, while making the transactions clear and transparent, generated large amount of digital data. Digital data has snowballed, with the proliferation of network services, mobile devices, and online users on the Internet the banking sector has witnessed a rise of interest in big data. A recent study showed that the global population produces a total of 2.5 quintillion bytes of data daily and this continue to rise. The global financial service sector continues to rapidly embrace digitization aided by the advent of new technologies and greater government push. This trend has been further stimulated by the emergence of Financial Technology (FinTech) players, who are playing a significant role in the banking value chain. The availability of Big Data in this domain has opened up new avenues for innovation and has offered huge opportunities for growth and sustainability. Simultaneously, it has presented a number of new challenges that must be surmounted in order to maximize its value. Since today's customers expect more personalized banking services and to remain competitive while complying with bigger regulatory surveillance, the banking services sector is hugely pressurized to utilize the depth and breadth of the available data in the best possible way. This study will examine big data and its analytics while looking at the impact and applications of Big Data in the banking domain. It will also consider the important Big Data challenges that remain to be addressed in the banking sector.

Keywords: Big Data, Big Data Analytics, Financial institution, Banking.

Introduction

As banking is becoming branchless, contemporary and digital at a very fast pace, the need to manage big data and analytics becomes more relevant as banks compete to gain competitive advantage. In recent years, as huge as 90 percent of the world's data has been developed due to the creation of 2.5 quintillion bytes of daily data. This speedy enlargement and storage of big data produce opportunities for structured and unstructured data collection, processing and analysis (Trevir, 2018).

Big Data Phenomenon, which is characterised by rapid growth of volume, variety, velocity, veracity and value of data - information assets, thrives the paradigm shift in analytical data processing. However, the extremely large financial data volumes, high generation speeds, and heterogeneity associated with the relevant financial domain data, along with its susceptibility to errors, make the ingestion, processing, and timely analysis of such vast volumes of often heterogeneous data very challenging (Taruna and Vipin, 2016).

Big data analytics has recently become the core driver of innovation in the banking sector. Various studies have shown areas like econometrics, investment analysis, risk assessment, fraud detection, customer interactions analysis, trading and behaviour modelling where big data is helping the banks to optimize their performance. According to Vladimir (2018), having witnessed a total investment in 2016 of about \$21 billion in big data analytics, the banking sector domain has become one of the leading consumers of big data services, and an ever-hungry market for Big data architects, solution and customized tools. The continuous acceptance of big data will certainly transform the landscape of banking services. The data collected in banks are so complex that it is outside the ability of any traditional data software tool to manage. Hence big data analytical tools are employed to solve this issue of managing, storing and analyzing large and complex data. Analytics transform data into

helpful insights thereby helping to improve the way banks budget, market products and plan for the future.

Despite the banking sector increasing acceptance of big data, there tends to be some significant but surmountable challenges in the field.

The big data phenomenon

Many authors have tried to define big data in many ways (first by Garnter in 2001, later updated in 2012). Big Data was originally defined by Gartner as “a high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight discovery, decision making and process optimisation”. Also Navint (2012) viewed it as “a term that refers to data sets or combinations of data sets whose size (volume), complexity (variability), and rate of growth (velocity) make them difficult to be captured, processed, managed or analysed by conventional technologies and tools, such as relational databases and desktop statistics or visualization packages, within the time necessary to make them useful”.

From these definitions, we can successfully view big data as being mainly characterised by “3V” words. The first “V” is volume, which is a pointer of the sheer amount of data produced and analysed compared to other traditional sources. The second “V” denotes Velocity which implies that data is generated at great speeds and Variety, the third “V” suggests that there are different forms of this emerging data which are different from traditional systems. However, there is the Veracity and Value are recently added as the fourth and fifth characteristics worthy of consideration.

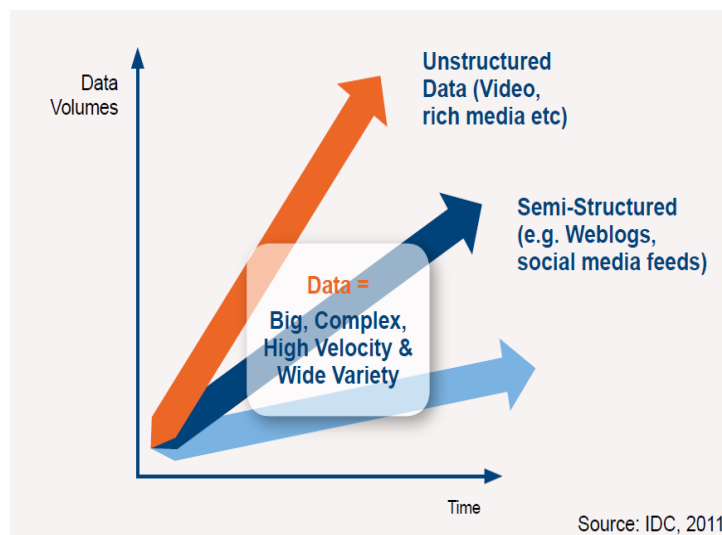


Figure 1. Definition of big data

Characteristic of big data

I. **Volume:** The quantity of analyzed data exceeds the capability of conventional analytics and statistical modelling tools. Big data itself relates to voluminous size hence size is germane to determining value out of data. Pence (2014) saw that the sheer volume of stored data is exploding, and IBM predicts that by 2020 there will be 35 zettabytes stored. The Centre for Economics and Business Research (2013) is of the view that as data volumes increase, there is a need for greater sophistication of technologies to harness the benefits that can be derived from it.

II. **Velocity:** This refers to the speed of accessing data. While conventional Business Intelligence (BI) applications use historical data dating back weeks, months or quarters, big-data relies on real time information in order to deliver rapid insights. According to Harvard Business Review (2012), the rate or speed by which data is produced is as important as the other characteristics as this provides rapid vision that can in turn provide competitive advantage and confirmed to be beneficial to organisations.

III. **Variety:** This represents information expansion resulting into heterogeneous data types (textual, numeric, etc.), encoding, formats, semantics, structures (structured, semi-structured,

unstructured), syntax etc. Usually, data can be separated into structured, semi-structured and unstructured data. According to Gartner (2013), unstructured data represents 85% of data while structured data is described as data grouped into rows and columns, making it easy to query and obtain information for an organisation's operational requirements. Semi-structured data, on the other hand, is a mixture of both structured and unstructured data.

IV. **Veracity:** This is tantamount to the quality of data. It involves the Origin, authenticity, availability, accountability and trustworthiness of data.

V. **Value:** This refers to the ability to convert a tsunami of data into business. Data about customers and requests has different value than data generated by GPS devices. Data can be transactional, descriptive, statistical, hypothetical or historical. Each of them has different value, hides different information (Andrej, 2013)

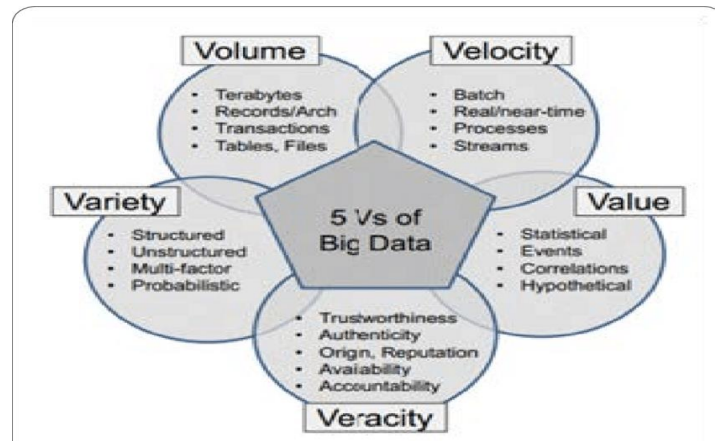


Figure 2. 5V's of Big Data

Big data analytics

The ultimate goal of a Big Data pipeline is to facilitate analytics on the available data. Big Data analytics provides the ability to infer actionable insights from massive amounts of data and can assist with information discovery during the process. It has become a core component that is being deployed and used by entities operating at various spheres of the financial field. For example, predictive analytics tools are increasingly being deployed by the banks to predict and prevent fraud in real time.

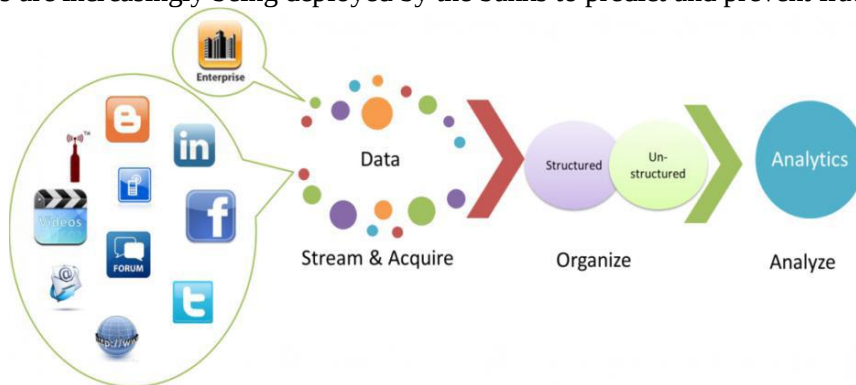


Figure 3. Data generation to analytics

To analyze a large volume of data, Big Data analytics is typically performed using specialized software tools and applications for predictive analytics, data mining, text mining, forecasting and data optimization (Vangie, 2018). Revealing the hidden information in big data via Data Mining (DM) techniques has become an emerging trend and ultimate objective for a wide range of studies (Hassani et al, 2018) Banking analytics can help improve how banks acquire, segment, target and retain customers. In addition, predictive analytics can help customers manage their accounts and easily complete financial transactions. Banks also benefit by minimizing risk and reducing costs.

Big data and analytics in banking

The mining of Big Data provides a massive opportunity to stand out from the competition as banks are now digitized. All banking transaction is a lump of data, so the industry sits on huge stores of information. By using data science to collect and analyze Big Data, banks can advance, or reinvent, practically every aspect of banking. Hence big data analytics has become a key driver of innovation in the financial sector. According to Ayesha and Riyazuddin (2017), data science can allow hyper-targeted marketing, personalized wealth management advice, optimized transaction processing and more - the prospect is endless. The application of big data analytics can also help support risk management in areas like credit scoring, cards fraud detection, financial crime compliance, stress-testing and cyber analytics. The recent desire by various governments' policy to digitize the entire banking sector and its subsidiaries greatly propels Big Data generation opening up huge opportunity globally. This available data shows that more than 1.7 billion mobile phone users are currently excluded from the formal financial system thus making them invisible to credit bureaus, but they are increasingly becoming discoverable through their mobile presence. This is boosting the global demonetisation to ensure cashless society.

While banks have traditionally been good at running analytics at a product level, such as credit cards, or personal loans, very few have done so holistically, looking across inter-connected customer relationships that could offer a business opportunity – for instance, when an individual customer works for, supplies or purchases from a company that is also a client of the bank. The dynamic data science field aids this seamless view.

Richa (2015) posited that less than half of the banks analyse customers' external data, such as social media activities and online behaviour. At the same time the use of the following is increasing:

- i. **Online Banking-** As the global usage of internet has increased to 56.1 percent as at April 2019, Internet banking facilities and mobile apps have consequently increased on a daily basis. Major attraction for the customers is the ease of paying utility bills online, money transfer, managing fixed deposit, buying insurance policies are attractive to the customers.
- ii. **Automated Teller Machine (ATM)** - The conveniences with which ATMs dispense money and accept deposits have become the biggest source of attraction to bank's customers. Also, increasing competition amongst banks is resulting into boost in ATMs installation trying to expand geographical areas. Government's cashless drive is playing crucial role in widening the sphere of ATM network focusing on rural India.
- iii. **Debit Card** - Banks are now issuing Debit Cards to their customers with saving or current accounts. These cards can be used by customers to buy goods and services at different channels like online, ATMs, Point of Sale (POS) etc instead of using cash. The amount paid through debit card is automatically debited (deducted) from the customers' account.
- iv. **Mobile Banking-** As the number of phone users increases, the use of mobile banking is becoming popular among both the educated and illiterate populace. Through the mobile phone, a customer can receive and send messages (SMS) from and to the bank in addition to all the functions possible like mobile top-ups, transaction viewing, sending and receiving cash in and out of their mobile wallets etc through phone banking.
- v. **Credit Card** -This is one of the products aggressively sold by banks to enhance the buying power of customers. It serves as a source of soft credit line on cards to help meet urgent needs of the customers.

Benefits of big data analytics to the banks

According to the study by IDC, the worldwide revenue for big data and business analytics solutions is expected to reach \$274.3 billion by 2022. Going by a recent Strategy Analytics survey of about 450 companies worldwide, banking tops most sectors when it comes to Big Data analytics. Big data that is either obtained from various sources or internally generated data is expected to be used in the manner that it is in harmony with the vision and mission of the organisation. Banks ensure they use this information to meet their predetermined objectives of either reducing cost, boost profit, minimizing the processing time, launch a new product to name but a few. These and others factors and variable should eventually result to better decision making in the organization. In as much as importance of big

data analytics to the banks is huge, the greatest prospect of big data in banking is still yet to be harnessed. The following are quite a few of the many examples of how big data is used in banking:

i. **Marketing and sales automation-** According to Peppard (2013), marketing analytics involves the practice of measuring, managing and analyzing market performance to maximize the effectiveness of and return on investment (ROI) from the marketing activities. Nowadays, with the huge volumes of available data, banks can gather previously unimaginable information about each of their customers. This provides a better understanding of customers' needs and helps to proactively address these needs. It also allows different departments within a bank, such as marketing, sales and IT, to collaborate for optimal performance. For instance, rather than pushing out products to all customers, banks can now merge Business Intelligence (BI) and sales force automation tools to market products tailored to customers' existing situations, whether they're constructing new homes, starting small businesses or beginning families. Indeed, banks using BI reported a 7 percent expansion in cross-sell and up-sell revenues

ii. **Product innovation, performance analytics and budgeting-** Banks employ BI to appraise business and employee performance and then create branch budgets and employee goals based on past achievements. Big data solutions allow banks to collect, analyse and share branch (as well as individual employee) performance metrics across departments in real time. This implies better visibility into the daily operations and a superior ability to proactively resolve any issues. Also, they can plan the education and training of their employees for off-peak times and monitor progress toward goals in real time. Banks can also use performance data on product features and services can be used by banks to produce new offerings designed around modern customer demand.

iii. **Risk Management and Fraud Prevention-**David and Desheng (2017) were of the opinion that there exist two instances of pioneering use of data analytics, machine learning and big data in banking institutions namely risk management and fraud prevention. In the area of risk management, big data analytics tools support bankers with deeper insights into their customer's behaviours by analyzing information like credit reports, spending habits and repayment rates of credit applicants. Big data software determines the chances that an individual would default on a loan or fail to constantly meet payment obligations as at when due (Hortonworks Inc, 2013). In preventing financial fraud, Urban (2014) established that big data analytics have become an essential part of any strategy to help detect and prevent financial crime, as a result of the ever-changing attack methods used by criminals exploiting multichannel vulnerabilities to compromise technology systems. Hence risk management and fraud prevention are two of the most important issues for banks at the moment and, for this reason, they are the first projects to have been addressed with these technologies. For instance, CitiBank strategically invested in Feedzai, a data science corporation that apply real-time machine learning and predictive modeling to explore big data to identify fraudulent behaviour and minimize financial risk for online banking providers.

iv. **Enhanced need assessments-**After getting access to huge amount of data, containing needs of different customers, banks can offer those needs in a meaningful way. By using this data, banking industry will provide exactly the information required by the customer instead of any other information.

v. **Personalized financial products-** Banks aspire to offer their customers a new account concept and tailor-made products by replacing the generic services with value-added ones. Since every customer has their own economic activity, the use of data analytics helps detect patterns and behaviours. This will enable banks to offer their customers personalized, bespoke financial products and services for an enhanced customer experience and better satisfaction.

vi. **Optimization of the bank's processes and resources** - Data gathering and analysis of bank's processes and resources helps them discover the pattern and behaviours that were, until now, unknown to maximize benefits while minimizing expenses. Consequently, the processes are significantly improved.

Challenges of big data and analytics

With the huge potential in big data management, it however comes with many challenges. It presents a number of challenges relating to its complexity. How big data can be understood and used

by banks when it comes in an unstructured format, such as text or video or how they can capture the most significant data as it occurs and deliver that to the right people in real-time or how they can store and analyze it given its size and our computational capacity. And there are numerous other challenges, from privacy and security to data quality and the likes which are considered thus:

- i. **Privacy and Security** - Big Data offers immense possibilities of providing major steps forward for Banks, but it is not without the challenge of privacy and intrusion. While the potential for abuse of this data is important, but Banks need to get it right. Thus, Big Data techniques and analytical tools can help customers get better service and assist Banks to effectively manage resources. It's a thin line between being helpful and intrusive (Mousumi, 2014).
- ii. **Data Quality** -It is essential that banks create Data Quality metadata that includes Data Quality attributes, measures, business rules, mappings, cleansing routines, data element profiles, and controls for Big Data Quality and Data Management. Since the greatest impact of big data is on its quality, data Quality is thus vital in this Big Data era. To guarantee highest Data Quality and Integrity, Data Quality attributes like validity, reasonableness, accuracy, completeness, timeliness and the likes must be clearly measured, defined, recorded, and provided to end users. Proper care must be taken not to lose the original values when data is mapped or cleansed. Data element profiles which ensure the completeness of every record must be created. Considering that data could migrate across systems, controls and reconciliation criteria have to be created and recorded to ensure that data sets accurately reflect the data at the point of development and that no duplicate or loss of data occur in the process. There is need to give special care to unstructured and semi-structured data as Data Quality attributes may not be easily or readily defined. If structured data is created from unstructured and semi-structured data, the creation process too must be documented and any of the previously noted Data Quality processes applied.
- iii. **Legal and Regulatory Challenges** - Big Data often come with great legal and regulatory concerns that have complexities and limitations due to sheer size. Many organisations already have control and data management procedures in place for small data, and a comfort level that those controls are appropriate. Considering the growing impacts of regulation and oversight, banks are avoiding Big Data, or at least proceeding judiciously, simply because of the risks.
- iv. **Talent Challenge**- In as much as good talent is scarce, getting that magic combination of hard science and business acumen is scarcer still. Blending a staff of data scientist skill and visualization teams is a new workforce management concept. Big Data Specialist need to have solid business understanding of SAS/R/SQL/Python programming and statistical knowledge along with Visualization skill. Also, Banks need to pay a lot more to a Data Scientist or a Big Data Specialist compared to a traditional ETL or Business Intelligence hire
- v. **Organizational Mindset** - Several Banks are still driven majorly by past experience, Intuition, SME knowledge and Customer Experience. There is need to have additional Data Curiosity, Data Driven thinking and to invest more in acquiring, storing and analyzing data. Banks that tap into Big Data and analytics will be more competitive, especially in this strict regulatory environment. Hence the need for banks to give data investment the top priority.
- vi. **Inefficient data management.**
- vii. **Regulatory Reporting.**
- viii. **Offer management and relationship pricing.**

Conclusion

The present information Technology age is characterised by generation of huge data. However, advanced analytics are making it possible for banks to manage the increasing cost of compliance and the risk of non-compliance. At present, big data analytics is mandated across numerous spheres of banking sector thereby helping them deliver higher services to both their internal and external customers, in addition to improving their security systems. Banks has realized that each of its customers generates massive amounts of data each and is fully reforming itself to extract as much hidden facts from this information as possible. Before now, this data did not add value but now the goal is to have the customer and their data at the nucleus of the business. This study focused mainly on big data and analytics as it relates to modern day banking. Besides showcasing the impact or

benefit of Big Data analytics in the banking, the challenges of the concept were critically reviewed. To remain competitively relevant, banks need to reconsider their *modus operandi* and adopt data-driven approaches. Moreso, the application of big data and analytics in the banking sector can help improve efficiency leading to customers' satisfaction and increased profitability.

References

- [1]. Andrej Kalicky (2013), High Performance Analytics, Master's thesis submitted to Department of Software Engineering, Charles University in Prague.
- [2]. Ayesha Anam Siddiqui and Dr. Riyazuddin Qureshi (2017), Big Data in Banking: Opportunities and Challenges Post Demonetisation in India, IOSR Journal of Computer Engineering (IOSR-JCE), e-ISSN: 2278-0661, p-ISSN: 2278-8727, PP 33-39. www.iosrjournals.org.
- [3]. Centre for Economic and Business Research (2018) <https://cebr.com>.
- [4]. David L. Olson and Desheng Wu (2017), Predictive Data Mining Models. Computational Risk Management, Springer, ISBN 978-981-10-2542-6, pp. 1-97
- [5]. Gartner (2013), Driving Value from Big Data (online) <http://www.gartner.com/technology/research/big-data/> [Accessed 09 May 2019].
- [6]. Harvard Business Review (2012), Big Data: The Management Revolution. Harvard Business School Publishing Corporation. Available: <http://automotivedigest.com/wp-content/uploads/2013/01/BigDataR1210Cf2.pdf> [Accessed 05 May 2019].
- [7]. Hassani, H, Huang, X and Silva, E. (2018) Digitalisation and Big Data Mining in Banking. Big Data Cogn. Comput. 2, pp 18.
- [8]. Hortonworks Inc. (2013), Banks looking for big data solutions to address credit risks. Retrieved from <http://hortonworks.com/big-data-insights/banks-looking-for-big-data-solutions-to-address-credit-risks/>.
- [9]. IDC (2019), <https://www.idc.com/getdoc.jsp?containerId=prUS44998419> (Accessed 02 June 2019).
- [10]. Mousumi Ghosh (2014), 5 Key Big Data Challenges in Banking Industry, Data Science Central (online). www.datasciencecentral.com/profiles/blogs/5-key-big-data-challenges-in-banking-industry.
- [11]. Navint (2012), Why is BIG Data Important? [Online]. Available: <http://www.navint.com/images/Big.Data.pdf> [Accessed 07 April 2018].
- [12]. Pence, H. E. (2014), What is Big Data and why is It Important? Journal of Educational Technology Systems, 43(2), 159-171. doi:10.2190/ET.43.2. d.
- [13]. Peppard, J (2013), The big data opportunity. Management Focus, pg 20.
- [14]. Philip Carter (2011), Big Data Analytics: Future architectures, Skills and roadmaps for the CIO, IDC White Paper.
- [15]. Richa Kapoor (2015), „Banking on Big Data”, Published on May 11, 2015, www.linkedin.com/pulse/banking-big-data-richa-kapoor?redirectFromSplash=true.
- [16]. Taruna Seth and Vipin Chaudhary (2016), Big data in Finance Trevir Nath (2018), How big data has changed finance, Investopedia. www.investopedia.com/articles/active-trading/040915/how-big-data-has-changed-finance.
- [17]. Urban, M. (2014), Big data analytics in the fight against financial crime. Retrieved from <http://www.bobsguide.com/guide/news/big-data-analytics-in-the-fight-against-financial-crime.html>.
- [18]. Vangie Beal (2018), Big Data Analytics, webopedia.
- [19]. https://www.webopedia.com/TERM/B/big_data_analytics.html.
- [20]. Vladimir Fedak (2018), Big Data Analytics in the banking sector, Data driven Investor (DDI) publication. www.medium.com/datadriveninvestors/big-data-analytics-in-the-banking-sector.