Association for Information Systems
AIS Electronic Library (AISeL)

SAIS 2014 Proceedings | Southern (SAIS)

4-14-2014

BIG DATA OPPORTUNITIES AND CHALLENGES: THE CASE OF BANKING INDUSTRY

Rudolph Bedeley
University of North Carolina, Greensboro, rtbedele@uncg.edu

Follow this and additional works at: http://aisel.aisnet.org/sais2014

Recommended Citation
Bedeley, Rudolph, "BIG DATA OPPORTUNITIES AND CHALLENGES: THE CASE OF BANKING INDUSTRY" (2014). *SAIS 2014 Proceedings*. 2.
http://aisel.aisnet.org/sais2014/2

This material is brought to you by the Southern (SAIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in SAIS 2014 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

BIG DATA OPPORTUNITIES AND CHALLENGES: THE CASE OF BANKING INDUSTRY

(Research in progress)

Rudolph T. Bedeley
University of North Carolina at Greensboro
rtbedele@uncg.edu

Lakshmi S. Iyer, Ph.D.
University of North Carolina at Greensboro
lsiyer@uncg.edu

ABSTRACT

The banking industry, with a large customer base and their use of mobile and other emerging technologies, has seen a surge in transactions leading to rapid generation of huge amount of data. This large amount of data presents great opportunities to the banking industry. At the same time, the industry faces huge challenges in managing the plethora technologies that are available to execute Big Data projects. Based on initial investigation, there is a gap in literature that clearly examines how the banking industry is leveraging the potentials of Big Data and challenges being encountered. Using a case study, this study seeks to fill the gap by investigating, at a more granular level, how the Banking industry is using and managing Big Data. Findings will contribute to knowledge and practice by increasing our understanding of Big Data implementation and management techniques from Banking Industry's perspective.

Keywords

Big Data, Banking, Industry, technologies, transactions, business intelligence, management.

INTRODUCTION

The term "Big Data" was first introduced to the computing world by Roger Mogoulas from O'Reily media in 2005 in order to define a great amount of data that traditional data management techniques cannot manage and process due to complexity and size (Ularu et al. 2012). However, a study of the Evolution of Big Data as a Research and Scientific Topic shows the term Big Data was present in research starting in the 1970s but has been censored in publications in 2008 (Halevi and Moed, 2012). Today, Big Data concept has permeated almost every industry and being treated from different points of view covering its implications in many fields.

Although there is not a concrete definition for Big Data, Manyika et al. (2011) refers to it as "data sets whose size is beyond ability of typical database software tools to capture, store, manage, and analyze." Big data has three main features: *volume* – huge data amount, *velocity* – speed of data creation, and *variety* – variety of source. However, experts have considered a fourth characteristic, *value* – providing valuable insights (DataStax 2011). Big Data can be seen in the finance and business environment where enormous amount of stock exchange, banking, online and onsite purchasing data flows through computerized systems every day and are then captured and stored for inventory monitoring, customer behavior and market behavior (Halevi and Moed, 2012).

During the past two decades, the banking industry has undergone radical changes, resulting in an environment characterized today by intense competition, globalization, heightened consumer mobility and demand, and deregulation (Cohen et al, 2007). But as most banks offer similar financial products and services (Cohen et al, 2007; Cengiz et al, 2007), they are revising their strategies from a product to consumer orientation to cope with the intense competition (Al-Hawari 2006). Bank managers realize that the challenge is not only to attract new customers but to retain and enhance relationships with existing ones (Iniesta and Sanchez, 2002). Achieving high levels of loyalty and recommendation (favorable word-of-mouth) can lead to larger market share, lower marketing and operational costs, and higher profitability (Ladhari et al., 2011).

As Barton and Court (2012) revealed in their study, data are essential, but performance improvements and competitive advantage arise from analytics models that allow managers to predict and optimize outcomes. This implies, however, that banks with the core capability to harnessing big data through daily transactions, customer-service records, real-time feeds, social media posts, and correspondence, have a competitive advantage at deriving more insights about their business over their rivals that lacks this unique capability.

Although big data comes with numerous opportunities and challenges across all industries, the following questions still remains unexplored and unanswered from literature: *(a) what is the nature of big data in the banking industry and what*

applications, approaches, and initiatives are being employed to handle big data management? (b) what challenges arise from the banks use of big data and what is being done to address these challenges?

This research is therefore designed to fill the gap by using a case study approach to explore, at a more granular level, the benefits banks are deriving from Big Data, challenges associated with its adoption, and strategies being implemented to address its issues. Findings from this study will not only help Bank Managers gain better insights about the antecedents and consequences of big data adoption, but also help them make informed decisions to the benefit their organizations.

LITERATURE REVIEW AND MOTIVATION

During the 1970s, information concepts like Management Information System (MIS) failed due the emerging information flood (Koreimann, 1971). Such an issue was addressed by approaches like the Data Warehouse (DWH) concept during the 1990s. But now, Big Data refers to a challenge of an unfavorable ratio between available data and current information technologies or concepts (Pospiech and Felden, 2012). According to Research Trends (2012) report, the concept of Big Data has received a heightened research attention lately in information systems (IS) as well as its sister research disciplines (e.g. marketing, computer science/engineering, etc.). A significant surge in Big Data research publications in top tier IS journals (MISQ, ISR, Management Science, JMIS, Decision Sciences, CACM, DSS, EJIS, etc.) has been phenomenal.

In order to obtain quality Big Data publications, we conducted an extensive search across top tier IS journals, marketing and computer science/engineering journals using key words: "Big Data, analytics, data mining, data warehouse" over the period 2005 – 2013. This yielded an initial 124 hits of publications. Top tier IS journals are those that occupy the top spot according to the MIS journal ranking due to the quality of their publications, high reputation, etc. The rationale behind limiting our search to only these journals is because of our interest in quality Big Data current research. To ensure that we have conducted an exhaustive quality literature search, however, other sister journals (e.g. marketing, computer science/engineering) were also explored. The articles were then filtered based on their relevance to Big Data application in business domain which eventually resulted in only 16 articles summarized and presented in Table 1. The table shows the study author(s), main findings, research method, and domain application of each study.

While these studies are not an exhaustive list of all the studies addressing Big Data issues, they represent a general overview of the most important research related to this study. The studies were chosen because their findings presents a generic idea about the origin of Big Data, current state-of-art, future direction, and potential challenges of its adoption. The main theme captured from the findings in Table 1 is challenges facing organizations' use of big data. Most of these studies were conducted through the use of a framework or conceptual model with the primary focus of providing generic guidelines/roadmap to enhance future research. Although only one of these studies (Chen 2011) talks about Big Data in financial markets, the scope is so broad that it does not specifically narrow down to the banking industry.

#	Reference	Summary of findings	Method	Application Domain
1	Jacobs (2009)	Addresses underlying technical issues of analyzing Big Data	Framework and conceptual model	Research
2	Panchaksharaiah (2009)	Shows relational database techniques to overcome Big Data	Framework and conceptual model	General IT practices
3	Agrawal et al. (2010)	Presents State-of-the-Art of Big Data and cloud computing	Survey and literature analysis	Cloud computing & DBMS
4	Eagle (2010)	Study present how complex human behavior data can be used to predict risky behavior	Framework and conceptual model	Predictive analytics / economic study
5	Bicking and Wimmer, (2010)	Text mining techniques get compared that support analysis in government applications	Qualitative and quantitative methods	e-Government, research theory improvement
6	Stonebraker et al., (2010)	Compares and quotes possible improvements of parallel databases and MapReduce	Framework and conceptual model	General IT practices

7	Cuzzocrea et. al. (2011)	Presents a State-of-the-Art of analytical Big Data technologies and cites current issues	Framework and conceptual model	Research
8	Agrawal et al. (2011)	State-of-the-Art of Big Data; Article	Framework and conceptual model	Research
9	Meijer (2011)	Introduce LINQ, a declarative, data-centric language that address Big Data	Framework and conceptual model	Research
10	Li et al. (2011)	Presents a DrayadLINQ, a declarative, data-centric language that addresses Big Data	Framework and conceptual model	Research
11	Reddi, et al. (2011)	Analyzes energy-efficiency of CPU approaches in Big Data environments	Field experiment	Research
12	(Bizer et al. (2011)	Discusses different challenges regarding to Big Data and semantic technologies	Literature analysis	Research
13	Alexander et al. (2011)	Discusses issues and changes of Big Data processing in scientific and engendering tasks	Literature analysis	Research
14	Satzger et al. (2011)	Shows issues of data management systems; Introduce a Big Data streaming technology	Framework and conceptual model	General IT practice & research
15	Chen (2011)	Big Data in the financial market to predict further stock markets based on Data Mining	Simulation / Mathematical analysis	General financial industry
16	Pospiech and Felden, (2012)	A State-of-the-Art framework about Big Data	Framework and conceptual model	Research

Table 1. Summary of Big Data Related Studies

According to Gartner (2012), 85 percent of all enterprise infrastructures will be overwhelmed until 2015 by Big Data. In this regard, Big Data related issues can be addressed from a functional as well as from a technical perspective. Here, traditional database management, storage, and analytical capabilities do not provide acceptable performance according to existing business requirements (Jacobs, 2009). Gartner's (2012) report shows a heat map of Big Data opportunities and challenges by industries. Figure 2 shows the five potential industries that would benefit the most, one of them being the Banking industry.

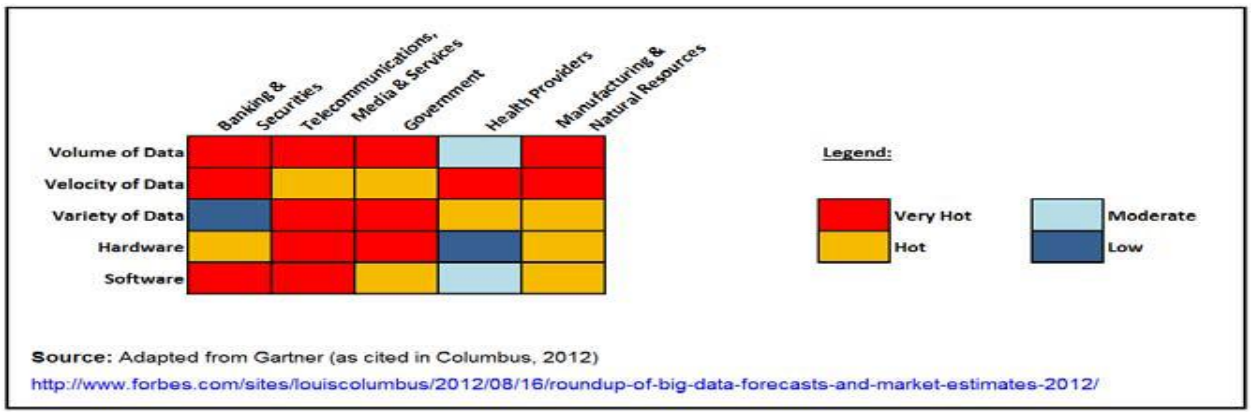

Figure 1. Big Data Opportunity Heat Map by Industry

The heat map presented by Figure 1 above shows that from the volume of data perspective, Banking and Securities, Communications, Media and Services, Government, and Manufacturing and Natural Resources have the greatest opportunity for Big Data. Likewise from velocity of data viewpoint, Banking and Securities still stand tall in regards to potential opportunity for Big Data. This result can be attributed to the huge amount of data that the banks generate daily from such a large pool of customers in their database. However, attention of variety of data currently being captured in the banking industry turns out to be low because, regardless of the means through which the data is captured, it usually is in structured format. This explains why Figure 1 shows banking as "low" in variety of data.

Data generated in the banking industry is massively growing at such an exponential rate due to several factors including hyper-competition, more business promotions to increase market share, increase customer base, large product sales, high innovation, etc. All these factors go into creating a huge volume of data which can be highly valuable if put into good use. Also, the advent of mobile banking and/or e-banking has increased the speed or velocity of data influx into the banking industry as a result of constant customer mobile interaction with their banks. Through the use of these technological gadgets, customers are able to remotely transact businesses with their banks such as wire transfers, online account balance check, making electronic bill pay, etc. As a result of the rapid generation of huge amount of data sets, the banking firms need to react to these exponential growth by upgrading both their software and hardware capabilities in order to take advantage of the opportunities from these data. This implies that their data capture, storage, processing, and analytic strategies and capabilities must significantly be improved in order to take up the looming challenge arising from their data surge.

Unfortunately, no prior study from our literature search seem to investigate, *at a granular level*, how the banking industries are leveraging the potential of Big data and how its associated challenges (volume, velocity, variety, and security) are being managed or addressed. Identifying this niche or research gap, we feel that there is a need for research that basically reveals the details of where the Banking industry actually stands as far as the opportunities and challenges associated with Big Data are concerned.

CHARACTERISTICS OF BIG DATA

There are three main attributes of Big Data. These are volume, variety, and velocity. Depending on the industry and organization, big data comprises of information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices.

Volume

In general terms, volume refers to the mass of quantities of data that organizations are trying to harness to improve decision-making across enterprise. Data volumes continues to increase exponentially as its sources continue to increase. The volume of Banks' data increases significantly from these possible source: consumer loans, demand deposit, commercial loans, mortgage loans, transaction files (ATM), credit desk, teller automation, etc. The increase in volume of data is attributed to significant increase in customer's demand for better products/service, increased competition, new technology, and many other factors. Currently banks are still using the conventional relational database management systems (DBMS) to store and manage their database. However, processing of their Big Data is being achieved through a more sophisticated technology such Hadoop and/or MapReduce.

Variety

Variety can be described as data of many formats which is compiled from a great deal of sources. These comprise of the Bank's transactional records or histories such as all the possible sources of both internal and external data, their Internet records (Facebook, Twitter, Blogs, Forums, etc.), business production or supply chain processes, geospatial datasets, and many more. Variety is mainly about managing the complexity of multiple data types, which are classified under three main categories: (i) *Structured data* – basically describes data which is grouped into a relational scheme (rows and column within a standard database), (ii) *Semi-structured data* – usually refers to data which possesses some form of relational structure or attributes, although it may not be complete or regular (e.g. XML files), and (iii) *Unstructured data* – usually refers to data in any format which cannot easily be indexed into relational tables for analysis or querying. Examples of such data type include text, web pages, social network content and blog posts, images, audio and video.

Velocity

This is defined as the speed at which data is created, processed and analyzed over time. The advent of e-banking or mobile banking technology and other sophisticate intelligent electronic devices has provided a remote access to customers where they frequently interact with their respective banks. Customers are always either checking their bank accounts, making online payments, transferring monies from one account to another, etc., all through Internet banking.

RESEARCH METHOD

This study uses a single case research method to address the research problem in a qualitative manner. Based on critical consideration of time and resource constraints, together with the fact that Big Data research in Banking Industry is limited and still evolving (Yin 2003), we believe the single case method is very appropriate to use. According to Yin (1994), several benefits can be derived from the use of case studies. For instance, the case studies enabled us to elicit subtle and rich data needed (in the form of explanations and descriptions) to address our research problem. Also, the open-ended interview approach used pointed us to obtain data from other sources (e.g. company archives, project reports, etc.). Finally, case studies enabled us to undertake an in-depth analysis of the application of Big Data in the Banking industry by answering "how" and "why" types of questions.

The Banking firm from which data was collected happens to be in its early stages of Big Data deployment with less than 3 years of experience. Data collected was administered through the use of interview protocol developed from consultation with both academic and practitioner literature. Interview content addressed what strategies the Banking Industry are using to harness the opportunities of Big Data and how possible challenges and issues are being addressed. The questionnaires on the interview protocol was reviewed and validated by experts with rich experience in Big Data. A sample of the final interview protocol is presented in Appendix I.

Data Collection

Semi-structured interviews was used to collect qualitative data from the head of IT in the bank that voluntarily participated in this study. With twenty years of experience in the field of analytics, the respondent has wealth of knowledge acquired through working with multiple banks of different sizes and geographical reach (local, regional, and national). Throughout these years he has served in different capacity in the IT department. He spent most of his time working as a business risk analyst, chief data officer, and has now risen through the rank-and-file to become the Enterprise Chief Data Officer. He has also served on several internal and external workshops and fora for financial services.

The interview followed a well-established procedure for conducting a semi-structured interviews as described in Paton (1990) paper. Within each of the topic area of the interview protocol, open-ended questions were typically asked before closed-ended questions so as not to bias respondent's answers. An initial open-ended format allowed both the interviewer and the respondent the opportunity to explore new leads and related topics. Answers to such questions are indicative of areas that are most important to respondents – and may serve to confirm or disconfirm our expectation. Also employed are standard probes, such as verification and compare-and-contrast questions, that helped elicit additional information.

With permission from respondent whose privacy of information was guaranteed through signed agreement document, information were gathered through the use of both hand-written notes as well as audio recording. Other supplemental data were gathered from secondary sources such as company publications, completed/ongoing project reports, etc. The interview was administered throughout the entire 50 minutes period by means of a long distance telephone discussion.

REFERENCES

1. Agrawal, D., Das, S. and Abbadi, A. (2011) Big Data and Cloud Computing: Current State and Future Opportunities, Proceedings of the EDBT, ACM, March 22-24, Sweden, 530-533
2. Alexander, F., Hoisie, A. and Szalay A. (2011) Big Data Guest Editors Introduction, Computing in Science & Engineering, IEEE, 2011, 10-13.
3. Al-Hawari, M. (2006). The effect of automated service quality on bank financial performance and the mediating role of customer retention. *Journal of Financial Services Marketing* 10 (3): 228 – 243.
4. Barton, D., Court, D., (2012). Making Advanced Analytics Work for You. Harvard Business Review, #79.

5. Bicking, M. and Wimmer, M. (2010) Need for Computer-Assisted Qualitative Data Analysis in the Strategic Planning ofE-Government Research, Proceedings of the Annual International Conference on Digital Government Research, ACM, May 17-20, Mexico, 153-162.
6. Bizer, C., Boncz, P. and Brodie, M. (2011). The Meaningful Use of Big Data: Four Perspectives, in SIGMOD, 40, 4, ACM, December 2011, USA, 56-60.
7. Cengiz, E., Ayyildiz, H. and Er, B. (2007). Effects of image and advertising efficiency on customer loyalty and antecedents of loyalty: Turkish banks sample. Banks and Bank Systems 2 (1): 56 – 78.
8. Cohen, D., Gan, C., Yong, H.H.A. and Chong, E. (2007) Customer retention by banks in New Zealand. *Banks and Bank Systems* 2 (1): 40 – 55.
9. Cuzzocrea, A., Song, Y. and Davis, K. (2011) Analytics over Large-Scale Multidimensional Data, Proceedings of the DOLAP, October 28, UK, ACM, 101-103.
10. DataStax (2011). *Big Data: Beyond the hype. Why Big Data matters to you*. DataStax, Burlingame, CA ,http://odbms.org/download/WP-DataStax-BigData.pdf
11. Denzin, N. K. (1978). The Research Act: A Theoretical Introduction to Sociological Methods, McGraw Hill, New York, 1978
12. Eagle, N. (2010) Big Data, Global Development, and Complex Social Systems, Proceedings of the SIGSOFT, ACM, November 7-11, USA, 3.
13. Gartner (2012). Market Trends: Big Data Opportunities in Vertical Industries. http://www.forbes.com/sites/louiscolumbus/2012/08/16/roundup-of-big-data-forecasts-and-market-estimates-2012/
14. Halevi, G., and Moed, H., (2012). "The Evolution of Big Data as a Research and Scientific Topic." Research Trends. Issue 30.
15. Iniesta, M.A. and Sanchez, M. (2002) Retail-consumer commitment and market segmentation . *International Review of Retail, Distribution and Consumer Research* 12 (3): 261 – 279.
16. Jacobs, A. (2009). "The Pathologies of Big Data." Communications of the ACM, 52, 8, ACM, August 2009, USA, 1-12.
17. Ladhari, R., Souiden, N., Ladhari, I. (2011). "Determinants of Loyalty and Recommendation: The Role of Perceived Service Quality, Emotional Satisfaction and Image." Journal of Financial Services Marketing. Vol 16, 2, pg. 111-124.
18. Li, H., Ruan, Y. and Qiu, J. (2011) Design Patterns for Scientific Applications in DryadLINQ, Proceedings of the DataCloud-SC, ACM, November 14, USA, 61-69.
19. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. and Byers, A.H. (2011). Big Data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation
20. Meijer, E. (2011) Big data is about more than size, and LINQ is more than up to the task, Communications of the ACM , 54, 10, ACM, October 2011, USA.
21. Panchaksharaiah, U. (2009). How to Adress Big Data, Communications of the ACM, 52, 12 ACM, December, 7.
22. Reddi, V. J., Lee, B. C. and Chilimbi, T. (2011) Mobile Processors for Energy-Efficient Web Search, ACM Transactions on Computer Systems, 29, 3, ACM, August 2011, USA.
23. Satzger, B., Hummer, W. and Leitner, P. (2011) Esc, Proceedings of the CLOUD, IEEE, July 4-9, USA, 348-355.
24. Stonebraker, M., Abadi, D. and Dewitt, D. (2010) MapReduce and Parallel, Communications of the ACM, 53, 1, ACM, January 2010, USA, 64-71.
25. Yin, R. K. (2003). Case Study Research: Design and Methods, 3rd Edition, Sage Publications, Thousand Oaks, CA, 2003.
26. Yin, R. K. (1989b). Case Study Research: Design and Methods, 2nd Edition, Sage Publications, Newbury Park, CA.