www.arpnjournals.com

BANKING ON BIG DATA: A CASE STUDY

Arti Chandani[1], Mita Mehta[1], B. Neeraja[2] and Om Prakash[3]
[1]Symbiosis Institute of Management Studies, Symbiosis International University, Pune, India
[2]MGR University, Maduravoyal, Chennai, India
[3]Real Apps Migrations, Pune, India

ABSTRACT

Big data; how big, is bigger than what the traditional application can handle and this gives a feel about the quantum of data which is being talked in the big data. Each day the technology is changing and everybody else is trying to cope up with the changes in the macro technological environment. Banks do generate a huge amount of data in their ordinary course of business which was being dumped in the books almost a decade back. Today the same data is being processed, analyzed and used for the benefits of the banks and customer. The data so generated can be used to customize services to the customer, to understand his needs, to design the most appealing marketing strategy to name a few. The big data, Peta-byte, can be efficiently used to analyze the financial behavior of a customer. A customer, who would have defaulted on a loan, may relocate making it difficult for the banks to trace but he still might be active on the social media, which can be used to trace the customer. This is one odd benefit which big data has to offer. All said and done, there are challenges to implement the big data technology for any bank. The biggest constraint comes from the finance front where any new technology requires a huge outlay of cash in the form of infrastructure, training and development cost and data warehouse and storage cost. The researchers have taken a hypothetical, yet practical, example to demonstrate the possible benefits of the adoption of the big data into a bank by calculating the net present value of the project. The researchers have used multiple rates instead of a single rate to help the users to take the net present value according to the rate applicable to them. The internal rate of return has also been calculated to understand the return which the project is generated itself and the same can be used by the users to compare with their internal rate of return to judge the viability of the project.

Keywords: big data, Indian banks, data storage, Hadoop.

1. INTRODUCTION

"Data is the new Oil. Data is just like crude. It's valuable, but if unrefined it cannot really be used"
-- Clive Humby, DunnHumby

We are living in the digital world and the data and the technology are integral part of the system. The technology has enabled us to use the transaction online while at the same time it has generated enormous amount of data which is somewhere eating up the storage space. At one side the technology is gearing up to provide more space and creating cloud technology to provide and meet up the requirement of the massive data which is being generated while at the same time others are busy in finding ways to use this data for their businesses and make it a business.

Big data is the data which is huge in quantity and which is captured by IT system in prevalence; it is too big and complex to be analyzed using the traditional software. The quantum and the speed at which data is being generated is tremendous; but, if analyzed and used in the right manner it could go a long way in benefitting the organization by offering them deep insight into a vantage situation thereby enabling a better decision making

The banks in the normal course of business generate huge amount of data and with the advent of IT and technology this data has grown multifold. The data or popularly called big data is being generated everyday with each financial transaction being carried out. The first known mover to have used the big data is HDFC bank which started using the big data in most efficient way and put in place a data warehouse and started investing in

technology that would help it make sense of the massive troves of unstructured data captured by its information technology (IT) systems. The more and more banks are gearing up to adapt and install ways and means to capture this data which should ultimately help them to improve their bottom line. The data so generated is to be analyzed in a way to provide insights into either the customer or the markets or both.

As commented by Munish Mittal, senior executive vice-president of IT at HDFC Bank, the stage was already set as early as in the year 2004, where the backbone of analytics was put in place. The enterprise data warehouse was already set up as a pioneering effort by the Bank. The motive behind the set up was to setup the capability to establish the differentiation of the customers based on the type of relationship they had with the Bank

The source of data for a bank could be many e.g. customer walk-in, emails, internet banking, voice call, social media, websites etc. The data analytics should be able to capture all the possible data and information for it to be used for the banks to analyse the customer, markets, and products. The data can be used e.g. to know whether you are the primary bank for the customers or not, what are different heads towards which the customers is spending money, and so on. This information can be used to provide solution to meet the requirement of the customer which might be very unique and could be otherwise known and offered.

2. BIG DATA

As the name would itself suggest, big data is an enormous or huge data-set, with a massive and complex volume so as to make it extremely difficult to process in the way traditional datasets are being managed as of today. The huge dataset pose excessive challenges in terms of analysing, capturing, storing, sharing, visualizing, presenting and securing, as it is unwieldy. To understand more on the nature of big data, it is often characterized by having:

1. Too much volume
2. Too much velocity (with which it comes or the speed with which it keeps coming to an organization such as a Bank)
3. Too much variety, as in today's context all sort of data with different formats, in the form of messages, images, clips, pdf, XML, etc, keep coming-up

Figure-1. Dimensions of Big data (Source: Palmer, 2013).

This is represented in the above figure. While there is huge variety of structured and unstructured data which comes up, the volume with which the new data is generated is also enormous. The volume is huge as the velocity with which the data comes is also huge. Today, everyone seems to be present in the virtual world of "online" activities, where everything seems to be done online. The Banking itself has become online, and one could hardly remember as to when was the last physical visit to Bank happened. The Virtual world of online activities has greatly expanded its domains. So whether it is airline booking, or cab booking, to shopping, paying taxes, buying gifts or just paying utility bills, everything has become online.

The way these online activities and transactions have increased off late, can be very well imagined by the fact that from the beginning till the year 2003, some 5 billion GBs of data was generated, as per one estimate. Same, amount of data (5 billion GBs) was created every 2 days in the year 2011! In 2013, same volume of data was generated in merely 10 minutes!! And this is an exponential acceleration. We definitely need Big Data.

Banks are no exception, where petabytes of data is getting easily generated. Most of the data is coming through the stupendous amount of online transactions,

authentications, authorizations, logs, audits, data mining, data analysis, backups, mirroring and so on. With the way data is accelerating, the traditional ways of managing the data is fast becoming obsolete.

The velocity is another dimension which creates enormous complexity to the scenario. Imagine that every 10 minutes, on an average some 5 billion GBs of data is arriving to be processed. What if, this data comes in few minutes or even seconds? How prepared is the organization to accommodate such data, is a Big Question being faced by the organizations today. The situation for the Banks is grim, as the financial data and applications are mission critical, and not even one transaction should be lost. Banks must be prepared to accommodate such Big Data at all costs.

The third dimension is the variety. Even if the Banks prepare for storing and processing the text data, like SWIFT messages, but what if the incoming data is in different format. What if it is an image format, an XML, PDF, XLS, DOC, JPG, MP3, MP4, AVI and so on. Such type of structured and unstructured data can be coming from various sources. For instance, sometimes the authentication can be based on the finger prints or other bio-metric data. Even such type of data has to be allowed. There are various cameras in the Banks premises, ATMs, and various other places. The data in the form of clips have to be stored. Thus, Big Data Technology is the need of this hour.

3. APPLICATION OF BIG DATA IN BANKS

The big data, either acquired from some source or internally generated data is to be used in the manner that is in sync with the organizational vision and mission. The banks should be able to use this data so as to meet the predetermined objectives which can be either to reduce cost, minimize the time taken in the processing, so launch a new product to name a few. All these and others factors and variable should ultimately lead to the better decision making in the organization

3.1 Advantages of Big data for banks

Using big data and technology, the banks may be able reap some of the following benefits:

- Find out the root cause of issue and failures
- Determine the most efficient channel for a particular customers
- Identify the most important and valuable customer
- Prevent the fraudulent behaviour
- Analyse the risk and the risk profiling
- Customised products and customised marketing communication
- Optimise human resources
- Customer retention

4. DATA ANALYSIS

A hypothetical example of a bank has been taken to illustrate the cost benefit analysis of the big data. The Net present Value (NPV) has also been calculated at

different rates to demonstrate the viability of the project. The cost of the traditional tools as well as for the big data tool for data analysis has been taken. The cost includes the hardware cost, Set up cost, Annual maintenance and support, Training and other costs. These costs have been taken assuming the 500TB data being generated every day by a medium size bank. The costs have been considered for a moderate period of 5 years which is assumed to be quite foreseeable from the strategic management view point. The training cost has been adjusted each year to arrive at a net figure of expected benefits. The cost includes the cost of the software which will include the application, data storage and data warehouse. The expected benefits of the big data would savings in the marketing budget, matching of product and customer to name a few. These benefits have been quantified to give a glimpse of the monetary benefits of the big data technology.

The expected benefits for both the tools have been analyzed by assigning the monetary benefits to the various variables. This has been done for example assuming that the traditional data analysis tool will bring in Rs. 2 customer savings and the number of customer are assumed to be 1, 20,000 for the first year. This gives the

monetary benefit for the first year as Rs. 25, 20, 000 while the cost for the first year are totaling to Rs. 21, 20, 000 brining in the net savings of Rs. 4, 00, 000 in the first year. This same procedure has been used for the remaining four years wherein the researchers have calculated the cost and benefits adjusting the value according to the various factors and the benefits. The benefits in the monetary terms are assumed to be increasing because of two reasons viz. increase in the number of customers and due to inflation the notional amount will increase and this has been duly captured in the data set.

The same technique has been used for the big data analysis tool by discounting the facts with respect to the cost and benefits. The big data tools are expected to be more efficient in terms of monetary benefits while the same is true when it come to the cost aspect of this tool. The cost, hardware cost is 3 times than the traditional tool cost which was assumed to be Rs. 20, 00, 000 while big data is assumed to be Rs. 60, 00, 000. The big data can bring in the benefits in financial terms which are equivalent to Rs. 6.5 in the first year and its ability to handle big data also get reflected in the number of customer being handled which were 1, 70, 000 in the first year.
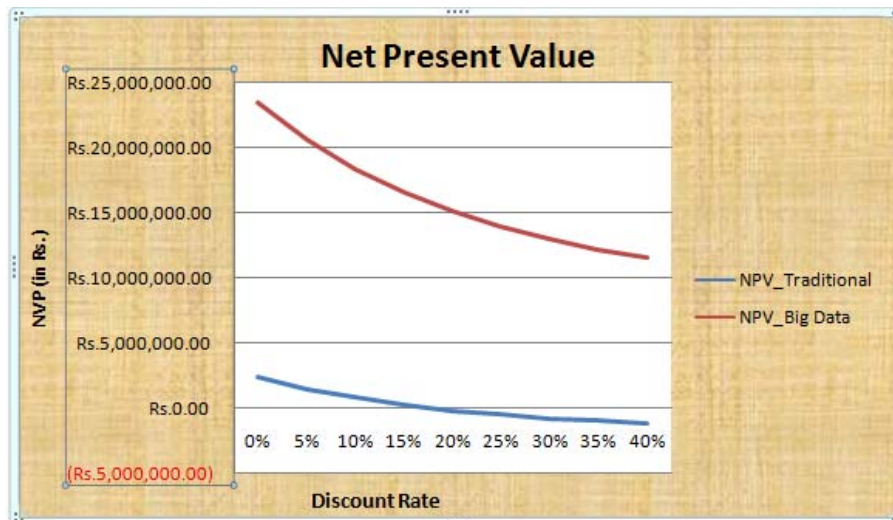

Figure-2. Net present value comparison for traditional vs big data.

The above figure shows the NPV for both the tools which clearly puts the case forward for big data despite of the higher initial cost. The researchers have calculated the NPV for both the tools at different rates so as to enable the users to apply the case as per convenience and applicability. The NPV of the traditional tool becomes negative at 20% while the same is positive at 15%, while the NPV of the big data is not only positive but much greater than the traditional tool across all the discount rates.

The researchers have also calculated the internal rate of return (IRR) for both the projects and it was found

that the IRR was 18.76% in the case of traditional tool while it was 32.50% in the case of big data tool. The internal rate of return shows the percentage return which the project is generating given the cost and benefits of the projects. This rate needs to be compared with rate applicable to the bank/enterprise to make a decision. If this rate is greater than the benchmark rate then the project should be selected and vice-versa.

www.arpnjournals.com

Table-1. Summary of the financial information.

Particulars	Traditional tool	Big data tool
NPV @ 10%	Rs. 8,57,484.28	Rs. 1,83,26,733.19
IRR	18.76%	32.50%

The above table shows the comparison between these tools over a period of 5 years and gives the values of most popular and widely used tools in the world of finance to enable the managers in decision making.

5. CHALLENGES OF BIG DATA FOR BANKS

Change is permanent but seldom anybody is ready to accept this and this is one of the most important aspects in the implementation of big data. The corporates and banks will have to make a gradual and swift shift towards "data culture" which may not be as easy as written on a piece of a paper. This not only calls for the overall strategic policy implementation but also the training and development of the employees to acquire the right skill set would be important parts of the big data programme.

The banks will have to identify the existing employee's current skill set and map the gaps required for implementation of the data analytics and cater for the various training programmes to address the issue. Furthermore the banks will have to align the recruitment policy for the big data and analytics to attract and retain the right talent. The banks will have to go the recruitment of the people who possess the skill set to handle and implement the latest technologies of the big data.

The implementation of big data calls for the technologies such as Apache Hadoop, NoSQL. This also calls for the investment in infrastructure which adds to the cost for the company. Infrastructure cost may not only include the physical assets but also the data storage and data warehouse is huge cost which calls for additional cost.

6. CONCLUSIONS

Big data is the reality and is going to stay there for a long time. It is important to note that the enterprises and banks are taking big data seriously as the banks haven cognizance of the fact that banking today is not more done in a protected environment and have to face the stiff competition not only from the public sector banks but also from the private and multinational banks. The banks needs to continuously adopt new technologies and system to remain ahead of the competition and big data is going to be a boon for this. Big data, no matter how big is the buzz word, but has its own set of limitations when it comes to the ground realities. Each bank will have to analyze its own policy for adaptation of the same weaving the organizational culture together as it is one of the most important of the whole process. The example taken here clearly demonstrates the monetary benefits which could be achieved by adapting the big data and the investment in

that would be a good move. On the contrary there will an annual cost with respect to the employee training and development. Moreover, there would be a strong need for data driven decision making rather than intuitive decision making, which is the bottom line of the big data.

REFERENCES

Richard Winter, Rick Gilbert, Judith R Davis. Big Data: 2014. What does it really cost? Wintercorp. Available from http://www.wintercorp.com/tcod-report/.

Anirban Sen. 2014. Banking on big Data analytics. Available from http://www.livemint.com/Industry/F5uNVbogJfsNB7cSt1toBL/Banking-on-Big-Data-analytics.html.

Bob Palmer. 2013. Getting the most out of big data and analytics. IBM White Paper. Available from https://www.ibm.com/smarterplanet/global/files/sweden_none_banking_mostoutofbigdata.pdf

Capgemini. 2013. Big Data Alchemy: How can Banks Maximize the Value of their Customer Data? Available from http://www.capgemini.com/resources/big-data-customer-analytics-in-banks.