

# ***Big Data Analytics for Banks***

*Bridging the gap between what is and what could be....*

## **Abstract**

BIG DATA ANALYTICS IS BIG NEWS! Data is growing at a tremendous rate with an increase in digital universe from 281 Exabyte's (in year 2007) to 8000 Exabyte's (in year 2015). The banking and financial services (BFS) industry has been one of the biggest adopters of Big Data technologies such as Hadoop. However, a lot of potential is unrealized when it comes to leveraging the data on Hadoop for analysis and decision making. One major issue that plagued the industry was mainframe data integration which required lots of resources with uncertain risks and rewards involved. Here, Veristorm Enterprise Solutions takes advantage of high-performance, streamlined data transfer solution which is easy to deploy and quicker than ever, and the performance gains achieved are remarkable. vStorm enterprise provides a single unified data integration platform to get all enterprise data on same platform where companies are storing external data. Leveraging vStorm Enterprise Solutions a bank can get a head start into using big data analytics for building out a recommendation engine, managing customer satisfaction and churn as well as managing fraud and credit risk.

## **Background**

As work started on writing this paper, big data analytics was one of the IT industry's hottest topics. A lot of the financial institutions / banks are currently shifting their marketing focus from traditional marketing focus that are outbound in nature to more point-of-impact real time fact-driven channel based marketing concepts. The traditional marketing concepts were mostly focused around communication channels like direct emails to reach out to prospect customer base but given that the world today is interconnected & so are the customers, banks can't just rely on only direct communication mode to interact with customers. To achieve customer profitability through product specific offers a traditional approach would earlier focus promoting a single product through appropriate marketing campaign wherein a group of customers showing similar behavioral pattern is expected to receive the same offer. Due to emerging trends & technology & customers shifting focus the banking industry is almost forced to re-invent the business model, develop new marketing intelligence & integrate risk management system to address some of key needs.

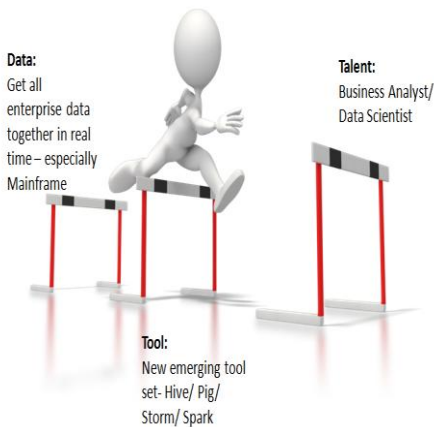
The external environment continues to be challenging for the Banking industry. The market is getting more and more competitive with entry of non-conventional players riding on wave of mobile payments. These new players are disrupting markets with their data analytics led innovative approaches. The customer loyalties continue to decrease with time as products have matured and there is reduced scope for differentiating ones products and services. Regulators across the world have become more active and are imposing more restrictive guidelines.

The obvious answer to key challenges above is for banks to leverage the available internal and external data and apply big data analytics. But despite the obvious answer why do traditional banks continue to

struggle to harness the big data challenge? We will explore the possible reasons and what steps banks could be taking starting today.

## The Challenge

The dynamics have changed rapidly over time and the current market state has created challenges & major pain points leading to a dilemma faced by C-level executives. The reasons for banks not realizing the full potential of available data, tools and technologies are not too hard to find out. One of the reasons is that all of the enterprise data is not available on the same platform and there continue to be silos across departments, product lines, geographies, etc. If one looks closer more often than not the historical transactional data from Mainframe systems continues to lag other sources in terms of availability and freshness. This is in part because historically the mainframe modernization projects have tended to be high cost involving uncertain risks and rewards. Of course through history of acquisitions or growth spurts banks also possess multiple systems. All the different systems add to the complexity of getting all data together on one platform.



some of them with steep learning curve. Whether the bank should go and hire resources who know Hive, Pig, Storm and Spark? Which skills should they focus on? What about the data that is stored in current data warehouses. Would bank be able to hire and retain these analytical resources? Which skills are really needed?



Let us examine this deeper. Hadoop based big data technologies have made it easier to ingest data and store on commodity hardware. So banks need not throw away any data. But what about historical data - How to get all the historical data from Mainframe onto Hadoop? While there are tools to import social media data onto Hadoop – What about a tool that is able to import recent data from Mainframe onto Hadoop in a click and point fashion.

Also, while getting data onto Hadoop is easy analyzing the same presents additional difficulties as traditional analytical tools either don't work with Hadoop or only limited features work on Hadoop. That opens up another front of learning new tools -

So do Banks have a long data migration project on their hand which would take months and years? Can they realistically use the mainframe data in making quick decisions on the fly? What options do the banks have in being able to harness the power of big data analytics? What kind of decisions bank could take if they could really harness the power of big data?

## The challenges require banks to refine their approach

The current environment can't only be treated as a technology shift – it has also changed the mentality of the customers too. The electronic media has exposed the customers to a whole new environment wherein electronic footprints are providing customers with varied new information almost every moment & opening up unlimited access to information bank almost about everything they need to make exercise their choice. In order to get a customer information, banks / organizations would need to understand and anticipate customer behavior in advance keeping in mind some of the key factors :-

- ✓ Outbound marketing offers through direct emails, calls may be annoying to the customer.
- ✓ Social media & mobile ecommerce have changed the dynamics empowering the customers look for better products in this interconnected world.
- ✓ Customers are looking for consistency across the communication they expect from the service – providers / banks.
- ✓ Customer needs may change over time with major life events – leaving university or school, getting married, purchasing a house, having a child, relocating or changing a job.

In order to address some of the factors banks need to be prompt in capturing the information at every point of time to provide improved customer experience consistently & enhance customer interactions with meaningful communication. One of the best ways to have such interactive communication is by taking the data of the customer from different sources & history and using Analytics to develop insights that enhances better targeting of customers.

*Example: Customer calls for settlement of personal loans.*

Traditional Approach: Customer to be redirected to the appropriate portfolio officer for settlement process which might take 2-5 minutes.

Event Based Approach: Customer transaction shows very high frequency of international trips. Offer immediate discounts on air-tickets purchased & special discounts on international purchases through bank credit/debit card that can be merged with continuity of personal loan.

The use of Analytics helps in revealing the hidden patterns of customer behavior & event-based marketing can be triggered to provoke the customers with product(s) matching their financial & lifestyle needs. The banks must as well need to be cautious about the policy or environmental changes that can impact customer loyalty & impact their decision making process. The factors to be kept in mind are:

- APR / interest rate changes
- Balance Transfer Schemes
- Introduction of competitive products at different price points

- Introductory offers on competitive price points

These events can be used to target & provoke a new customer from switching to a new portfolio / product or use a special personalized offer to address retention of high risk customers. Analytics enables banks & financial institutions to mine the customer's data (behavior, transaction, social, geo location etc) to be able to track customer's activity real time & use some of the insights gathered from the past to identify set of customers who needs to be addressed for acquisition or retention.

Summarizing the points discussed above the broad sense that is coming out from the discussion can be categorized under three (3) broad categories:

- ✓ Aggregate & Analyze data at rest & data in motion
- ✓ Predictive Modeling platform – transforming information into insights
- ✓ Decision system – Recommender Engine

### Banks best friends – vStorm and Spark

A number of companies have been researching ways to help enterprise address the above challenges. It is not that solutions do not exist. Hadoop has helped banks significantly in addressing data storage issue. But most of traditionally available tools and solutions suffer from the same issue. They require long projects running into months and years for implementation cycle. They need resources with specific IT skills and rely heavily on them. The technical nature of these tools means that average business user feels intimidated during implementation and is not fully engaged. Then as banks slowly get acquainted with the technology, new ones come up on the horizon. We at MathLogic, have scanned the market and two complimentary tool set seems to make sense. First one is a GUI based point and click tool to move data from **Mainframe** or any other **open systems like Oracle, DB2,SQL Server, Teradata** and other disparate sources into Hadoop from Veristorm (vStorm Enterprise). The second is the tool set for analysis from UC Berkeley which is fast gaining acceptance as the universal tool for analysis (Spark – in particular we think Spark over Hadoop would be the game changer for banking Industry).

#### **The vStorm Advantage**

As with any enterprise technology standard, Veristorm's big data integration tools is able to work with a variety of heterogeneous big data languages and sources while abstracting the user base from the complexity of implementation. The point and click user interface for moving Mainframe data is something which can finally leave the power in the hands of analyst without dependence on Mainframe IT resources. Finally, it should be noted that vStorm Enterprise can send data to external sources for processing. Using vStorm Enterprise, distributed Hadoop configurations can work with Cloudera, Hortonworks, and BigInsights platforms & organizations like IBM are using Infosphere applications along with vStorm as the platform of choice. Thus it makes the Analytics procedure much easier to get a holistic view of the customer data. A single view of the customer is an absolute necessity in today's world to be able to have the right analytics models built & implemented.

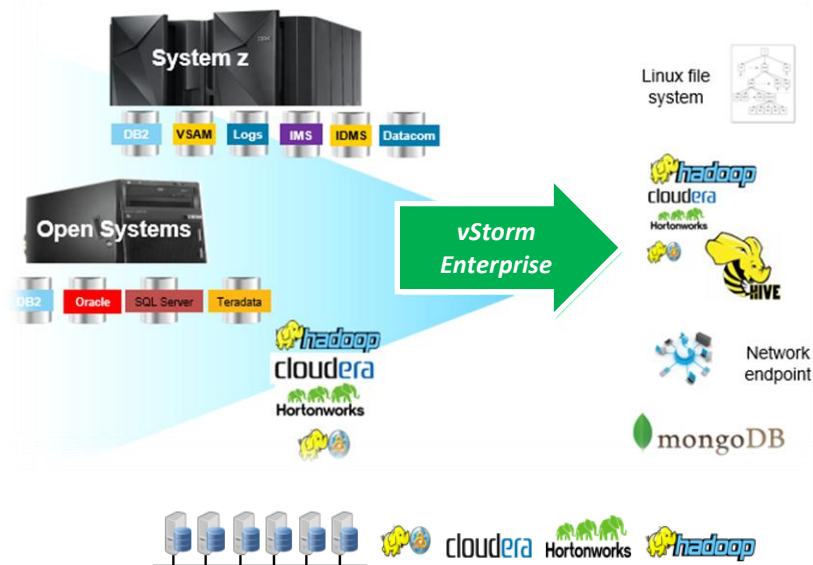


Figure 1: Leverage vStorm Enterprise

### The Spark Advantage

A number of tools have been promising to be the universal tool in the new age environment. Years ago SAS was ubiquitous amongst banks. Its ability to manipulate data as well as perform complex analytics need by bank almost made it mandatory for banks to have some kind of SAS license. Then slowly 'R' emerged as a more affordable analysis tool which could do most of what SAS did and more. It also offered promise of connecting seamlessly to Hadoop – something SAS was slow to pick on. Now it has emerged although R can be run in distributed mode not all functionality is available in R-Hadoop configuration. Banks either have already moved lot of enterprise data to Hadoop or are in the process of doing so. Being able to analyse the data effectively in Hadoop is crucial now. Here is where Spark can prove to be the difference. Conceptualized in UC, Berkeley it is now the most popular open source projects and promised to revolutionize analytical tool industry. It seems one of the most unified tools with components for running Machine Learning Algorithms along with SQL and streaming data directly utilizing data available on Hadoop clusters.

### vStorm - Spark combine as Predictive Analytics platform

A closer look at Figure 1 shows that data from historical process is gathered and collated into Hadoop platform through Mainframe systems which stores all the historical information gathered from the customer. Once the data from external sources [Social media, mobile ecommerce, credit bureau, etc] are also plugged into this system & clubbed together, we have solved the most of our problem of consolidated data view. Basis some of the key challenges the banks are facing today, the analysts or data scientists can now build robust predictive analytics models that can solve the business puzzle and enhance business value. They can use Spark as the single tool to analyze structured/ unstructured and streaming data on the go.

For example - the analytics models / decision rules can together be hosted on Recommender Engine platform & based on customer's real time activity each customer score / segment profile can be identified real time to provide him with the offer.

The information management platform builds on the Apache Hadoop project and incorporates latest technology from Veristorm. This address the first two hurdles in making Big Data Analytics work for a bank. Along with the environment it is almost a necessary condition in today's volatile environment to tap into the right set of resource possessing both analytical skills & banking domain expertise to solve some of the critical puzzles in the industry. Analytics / consulting organizations having the exposure to work on Hadoop environment & have familiarity with tools like Spark & Hive should be considered as a partner of choice for Banks. This can potentially be an ideal technology handshake between **Advanced Analytics & Information management**.

## Use cases for Banks

In a highly competitive market is driving firms to compete aggressively for customers' wallet: increasing focus on customer acquisition, retention and profitability. While getting a complete view of customer relationship across the enterprise is very important, it is equally essential to use it to offer customized products and service to profitable clients will increase client loyalty and result in increased wallet share & reduce losses by minimizing the risk exposure.

**Real time Recommended Offer:** Target new product & services to the right customers by implementing Analytics engine that supports flexible & integrated processes by better understanding customer needs, preference, buying patterns & motivations. The recommended offer campaign is meant to provide the customer a complete new personalized experience by communicating consistent message across channels to increase customer delight through each interaction. Instant access to customer activity real-time & analyzing customer profile from past history can be really helpfully in triggering an offer to the customer commensurate with his behavior, current location & accessibility to exercise that offer increases customer satisfaction to a huge extent.

Potential Benefits: - Enhanced customer satisfaction, increase customer-value & engagement level, personalized touch, Increased revenue from existing customer base

**Churn Management:-** As a business, Banks or Financial institutions typically start understanding the reasons for customer churn when they start analyzing the data across different portfolio & are able to sort some of the service related issues. The major threats in the banking industry are

- Not able to understand who will churn & when this might happen
- Challenges in understanding customer dissatisfaction by exploring historical information across different data sources.

In Today's world, the basis of customer relationship has changed drastically as the customers are always inter-connected, have access to multiple online devices & have more information towards competitive

products / offers. So customers today are well informed about what is happening around & the expectation levels of the customers are reasonably high. Analytics models & strategies can find out hidden patterns of customer's historical information & current needs / preferences to generate actionable insights on each & every individual. Models can also rank them from high to low risk with respect to attrition. Appropriate decisions & recommended actions can be triggered based on customer's propensity to churn to address overall portfolio balance & revenue. Win Back campaigns can as well be planned by properly addressing customer dissatisfaction level to get back recently churned customers.

Potential Benefits: - Understand customer dissatisfaction level, reduce portfolio churn, maintain portfolio profitability, identify major drivers for churn

**Fraud Management:** While fraud isn't a new problem for banking and financial services companies, many are struggling to find solutions that can prevent with the emerging trends & technologies of online purchase & more social media influence. The challenge is to maintain a positive, customized, and seamless customer experience by arresting fraudulent transactions real time to reduce false positives for case management.

The key to accurate and non-disruptive fraud detection is to implement emerging technology that allows banks to gain a holistic view of customers. Moving forward, with advancement of technology & analytics strategies, having security system can help detect online threats real-time can protect banks from damage losses.

Potential Benefits: - Reduce portfolio loss, increase bank reputation, protect customers, distinguish normal activity from fraudulent activity

## Summary

Understanding and utilizing the power of big data analytics has become imperative for retail banks. Potential applications for an effective analytics program are nearly boundless. During implementation, banks should adopt a structured approach and ensure that regulatory and compliance requirements are strictly met, proper stakeholders are involved. New applications are continually emerging in the market for big data Hadoop environments. With Veristorm products, mainframe data can be shipped to a different region on the mainframe (the Integrated Facility for Linux) region where it can be analyzed. The mainframe processor has never been more capable for processing analytic workloads. In the end, it all comes down to this – if the data resides in mainframe databases, that data should be processed by the mainframe. Veristorm's extract-Hadoop-transform approach enables certain stack-oriented Hadoop workloads to be processed on the mainframe – offering enterprises an efficient way to process structured and unstructured data on the most secure system in the industry. Adding on the tool set within Spark now makes it possible to analyse all the available data on the go and make real time recommendations. Those that do adopt this approach into firms' big data policy DNA structure, will achieve a significant competitive advantage – selling the right products to right customers at a right point of time.

## Call to Action

At a given point of time, when firms have the right vision and goals, they work as a team, and hold tools in a loving embrace that can shine a light well into the future, the prospects around big data is astronomical. With people continuing to see the myriad of possibilities, they will figure out ever creative ways to leverage the data IT makes available. To find out more about how your business can gain in smoothing your operations and maintaining structured channels across key businesses with FN MathLogic Consulting Services, do reach out to us at [info@fnmathlogic.com](mailto:info@fnmathlogic.com) ([www.fnmathlogic.com](http://www.fnmathlogic.com)).