Volume 3, Issue 12, December-2016, pp. 632-638 ISSN (O): 2349-7084

International Journal of Computer Engineering In Research Trends

Available online at: www.ijcert.org

A Survey of Big Data Analytics in Banking and Health Care today

[1]P.Siva Kumar, [2]B.Murthujavali,[3]Surya Pogu Jayanna

[1,2,3] Asst.Professor,Department of Computer Science and Engineering.
G.Pullaiah College of Engineering & Technology (GPCET), Kurnool.

Abstract: This paper gives an perception of how we can disclose added value from the data generated by healthcare and banking sector where Health care Organizations and Financial institutions are making use of Big Data in big ways, from boosting cyber security to reducing customer or Patient churn, cultivating customer or Patient loyalty, and more through innovative and personalized offerings that make effective services. Large amount of heterogeneous data is generated by these organizations. But without proper data analytics methods these data became useless. Big Data Analytics using Hadoop plays an effective role in performing meaningful real-time analysis on the huge volume of data and able to predict the emergency situations before it happens. It describes about the big data use cases in healthcare and banking sectors.

Keywords: Big Data, Hadoop, Healthcare,Banking, Map-Reduce.

I.INTRODUCTION

Big data analytics offers promise in many business sectors, and health care is looking at big data to provide answers to many age-related issues, particularly dementia and chronic disease management. This systematic review explores the depth of big data analytics since 2010 and identifies both challenges and opportunities associated with big data in health care. The review follows the standard set by Preferred Reporting Items for Systematic Reviews and Meta-analysis (2009) [1].

Big data is commonly defined through the 4 Vs: volume (scale or quantity of data), velocity (speed and analysis of real-time or near-real-time data), variety (different forms of data, often from disparate data sources), and veracity (quality assurance of the data). The first 3 Vs are found in most literature [2,3], and the fourth V is a goal [4].. In today's digital world, Health information systems show great potential in improving the efficiency in the delivery of care, a reduction in overall costs to the health care system, as well as a marked increase in patient outcomes [8]., it's necessary that large volume of data generated should be analysed effectively to answer new challenges. Similarly Financial Institutions also generates petabytes of data every day. It requires a technology that helps to perform a real time analysis on the enormous data set. This will help the banking to provide value added services to the citizens. Big data analytics helps in discovering valuable decisions by understanding the data patterns and the relationship between them with the help of machine learning algorithms (1). This paper provides an overview of big data analytics in healthcare and Banking systems. It describes about big data generated by these systems, data characteristics,

 DOI:10.22362/ijcert/2016/v3/i12/xxxx

security issues in handling big data and how big data analytics helps to gain a meaningful insight on these data set.

2. BIG DATA USE CASES

Big data in health-care refers to the patient care data such as physician notes, Lab reports, X-Ray reports, case history, diet regime, list of doctors and nurses in a particular hospital, national health register data, medicine and surgical instruments expiry date identification based on RFID data. Healthcare organizations are depending on big data technology to capture all of this information about a patient to get a more complete view for insight into care coordination and outcomes-based reimbursement models, health management, and patient engagement.

2.1 Prerequisite Big Data Analytics in Healthcare System:

Providing patient centric services: Increasing the use of technology is slowly changing the direction of the health care sector from disease-centric care toward patient-centric care [5]. Big data will play a significant role in this transformation. It will allow the information to be delivered to patients directly and empower them to play an active part in their care. When patients are provided with the appropriate information, it will influence their decision making and allow them to make informed decisions. Informed decisions will also be influenced by increased communication between patients, providers, as well as their communities. About 29% of the literature mentioned this opportunity. To provide faster relief to the patients by providing evidence based medicine-- detecting diseases at the earlier stages based on the clinical data available, minimizing drug doses to avoid side effect and providing efficient medicine based on genetic makeups(1). This helps in reducing readmission rates thereby reducing cost for the patients. Detecting spreading diseases earlier: Predicting the viral diseases earlier before spreading based on the live analysis. This can be identified by analyzing the social logs of the patients suffering from a disease in a particular geo-location (1). This helps the healthcare professionals to advise the victims by taking necessary preventive measures.

Improve Quality of Care

Big data has the potential and ability to improve the quality and efficiency of care. Big data offers an ability to predict outcomes using the available primary or historical data and provide proof of benefit that could change established, industry-wide standards of care. Leveraging technology at the patient end can also help with medication adherence. This will most certainly play an important role in improving outcomes and Globalization [11]

Globalization

Big data will actively help in disseminating the knowledge acquired from the data collected. Big data plays an active role in leveraging the practices and knowledge not only regionally but globally. By globalizing data, it is made more widely accessible and providers may access new information from all regions. About 24% of the literature mentioned this opportunity [11].

Fraud Detection

One of the most significant benefits offered by big data is that it is instrumental in detecting fraud in an efficient and effective manner. For example, the unauthorized use of specific user accounts by third parties can be minimized. Only about 11% of the literature mentioned this opportunity.ove the health-related quality of life. Quality of care will also be improved by reducing waste of information, which will reduce inefficiencies [4]. This will also assist in analyzing real-time resource utilization productivity [4]. Quality can also be improved by reducing the rates of readmissions, increasing operational

 DOI:10.22362/ijcert/2016/v3/i12/xxxx

efficiencies, and improving performance. About 64% of the literature mentioned this opportunity[11].

Monitoring the hospital's quality: Monitoring whether the hospitals are setup according to the norms setup by Indian medical council. This periodical check-up helps government in taking necessary measures against disqualifying hospitals. Improving the treatment methods: Customized patient treatment---monitoring the effect of medication continuously and based on the analysis dosages of medications can be changed for faster relief. Monitoring patient vital signs to provide proactive care to patients. Making an analysis on the data generated by the patients who already suffered from the same symptoms, helps doctor to provide effective medicines to new patients.

2.2 Big Data Analytics in Banking System:

With the Big Data and analytics market predicted to reach $125 billion in 2015 (IDC), nowhere is this having greater impact than the finance industry. Banks and financial institutions generate vast volumes of data, with numbers accelerating exponentially every day. In this quick-fire market, high-pressure decisions have to be made rapidly, guided by a volume of information that simply did not exist just a few years ago.

The explosion of data is driving financial institutions to leverage advanced analytics to tackle a myriad of challenges and gain commercial advantage. By unlocking the wealth of information contained in market feeds, customer-service records and social media data, banks can derive more insight about their business than ever before, and shape future strategy through factual insight rather than intuition.

Addressing advantages with big Data analytics

Fraud Detection & Prevention

One of the biggest problems faced by the banking sector is fraud. And Big Data will allow banks to make sure that no unauthorized transactions will be made, providing a level of safety and security that will raise the security standard of the entire industry.

Enhanced Compliance Reporting

Banks now have access millions or even billions of customers' needs, and they can now use Big Data to cater to them in a more meaningful way. Cloud based analytics packages can sync in real time with your big data systems, creating actionable insight dynamically. Big Data will expand the banking industry in a way that will allow them to earn mores revenue through cost reduction. And by cutting down on unnecessary costs, the banking industry can provide customers with exactly what they're looking for, instead of irrelevant information.

Customer Segmentation

Big Data will give banks deep insights into customer spending habits and patterns, simplifying the task of ascertaining their needs and wants. By being able to track and trace each and every customer transaction, banks will be able to categorize their clients based on various parameters, including commonly accessed services, preferred credit card expenditures, or even net worth. The benefit of customer segmentation is that it allows banks to better target their clients with relatable marketing campaigns that are tailored to cater to their requirements. Herman Shooster, Founder of Globalresponse.com, was a big advocate of Big Data for Customer segmentation.

Personalized Product Offerings

Customer segmentation can further be used to create and deliver new schemes and plans, aimed directly at the specific requirements of their customers. By analysing past and present expenses and transactions, a bank can get a clear understanding of how to get the highest response rate from their clients. Creation of personalized product offerings will cater to an

 DOI:10.22362/ijcert/2016/v3/i12/xxxx

untapped niche of personalized services that gives banks the ability to create more meaningful client relationships.

Risk Management

The early detection of fraud is a large part of risk management, and Big Data can do as much for risk management, as it does for fraud identification. Big Data locates and presents big data on a single large scale that makes it easier to reduce the number of risks to a manageable number. Big Data plays a pivotal role in integrating the banks requirements into a centralized, functional platform. This reduces the banks chances of losing data, or ignoring fraud.

By keeping up with Big Data and other newer global trends, the banking industry will be able to get a better understanding of client requirements, so that they can provide such services in a timely manner. The task of implementing Big Data on a large scale is just taking shape, with many IT departments concerned about the transition to high-tech IT infrastructure. But when it comes to Big Data adoption in the banking industry, the sooner the better!

3. IMPLEMENTATION OF BIG DATA ECOSYSTEM

The Secured Big Data architecture of healthcare is shown in figure 1. Electronic health record is a heterogeneous data set which is given as input to HDFS through flume and sqoop. Analysis on the data is performed using Map-Reduce and HIVE by implementing machine-learning algorithms which helps in analysing similar pattern of data. This helps in predicting the risk of patient health condition at the earlier stages. Hbase is used for storing the multi-structured data. STORM is used to perform live streaming and any emergency conditions such as patient temperature rate falling beyond the expected level can be intimated to care-takers immediately through AWS Lambda function. Report is generated through intellicius and hunk.

3.1 Big Data Ecosystem for Healthcare and Government

It is a complex system that constitutes of components and technologies to handle large scale data processing and analytics on it. It includes getting the data from various sources, store them in HDFS (Hadoop Distributed File System), process the data using Hadoop components such as Map-Reduce, perform analysis using PIG and generate Business Intelligence reports such as patient score cards.

3.3 Big Data Lifecycle

3.3.1 Data Collection: It involves the collection of data from various sources and storing it in HDFS. Data can be anything such as case history, medical images, social logs, sensor data etc.

3.3.2 Data Cleaning: It involves the process of verifying whether there is any junk data or any data that has missed values. Such data needs to be removed.

3.3.3 Data Classification: It involves the filtering of data based on their structure. For example Medical Big data consists of mostly unstructured data such as hand written physician notes. Structured, semi-structured and unstructured data should be classified in order to perform meaningful analysis.

3.3.4 Data Modeling: It involves performing analysis on the classified data. For example Banking may require the list of malnourished children in a particular location (8). First it has to classify the data based on the specific location, need to trigger the health report of children, need to identify the children whose families are under poverty line and these data should be processed.

3.3.5 Data Delivery: It involves the generation of report based on the data modeling done. Based on the example after the data is processed it will generate a

report based on malnourished children in a particular location. This will help the government to take necessary measures to avoid any further complications. At the all the stages of BDLC (Big Data Lifecycle) it requires data storage, data integrity and data access control Big Data Every day one hears more about how Big Data ecosystem technologies are helping create incremental innovation & disruption in any given industry vertical – be it in exciting new cross industry areas like Internet Of Things (IoT) or in reasonably staid areas like Banking, Manufacturing & Healthcare. Big Data platforms, powered by Open Source Hadoop, can economically store large volumes of structured, unstructured or semi structured data & help process it at scale thus enabling predictive and actionable intelligence.

Corporate IT organizations in the financial industry have been tackling data challenges at scale for many years now.

Traditional sources of data in banking include

1. Customer Account data e.g. Names, Demographics, Linked Accounts etc
2. Transaction Data which captures the low level details of every transaction (e.g debit, credit, transfer, credit card usage etc),
3. Wire & Payment Data,
4. Trade & Position Data,
5. General Ledger Data and Data from other systems supporting core banking functions.

Shortly after these "systems of record" became established, enterprise data warehouse (EDW) based architectures began to proliferate with the intention of mining the trove of real world data that Banks possess with an intention of providing Business Intelligence (BI) capabilities across a range of use cases – Risk Reporting, Customer Behavior, Trade Lifecycle, Compliance Reporting etc. Added to all of this, data architecture groups are responsible for maintaining an ever growing hodgepodge of business systems for customer metrics, adhoc analysis, massive scale log processing across a variety of business functions. All of the above data types have to be extensively processed before being adapted for analytic reasoning.

You also have a proliferation of data providers who want to now provide financial data as a product. These offerings range from Market Data (e.g. Bloomberg, Thomson Reuters) to Corporate Data to Macroeconomic Data (e.g Credit Bureaus) to Credit Risk Data. Providers in this business like the above typically construct models (e.g credit risk) on top of these sources and sell the models as well as the raw data to interested parties. Thus architectures need to adapt in an agile manner to able to scale, ingest and process these feeds in a manner that the business can leverage to react to rapidly changing business conditions. Thus, Bank IT world was a world of silos till the Hadoop led disruption happened.

4. HADOOP FOR FINANCIAL SERVICES

The key challenges with current architectures in ingesting & processing above kinds of data –

1. A high degree of Data is duplicated from system to system leading to multiple inconsistencies at the summary as well as transaction levels. Because different groups perform different risk reporting functions (e.g Credit and Market Risk) – the feeds, the ingestion, the calculators end up being duplicated as well.

2. Traditional Banking algorithms cannot scale with this explosion of data as well as the heterogeneity inherent in reporting across areas such as Risk management. E.g certain kinds of Credit Risk need access to around 200 days of historical data where one is looking at the probability of the counter-party defaulting & to obtain a statistical measure of the same. The latter are highly computationally intensive.

 DOI:10.22362/ijcert/2016/v3/i12/xxxx

Circa 2015, Open source software offerings have immensely matured with compelling functionality in terms of data processing, deployment scalability, much lower cost & support for enterprise data governance. Hadoop,

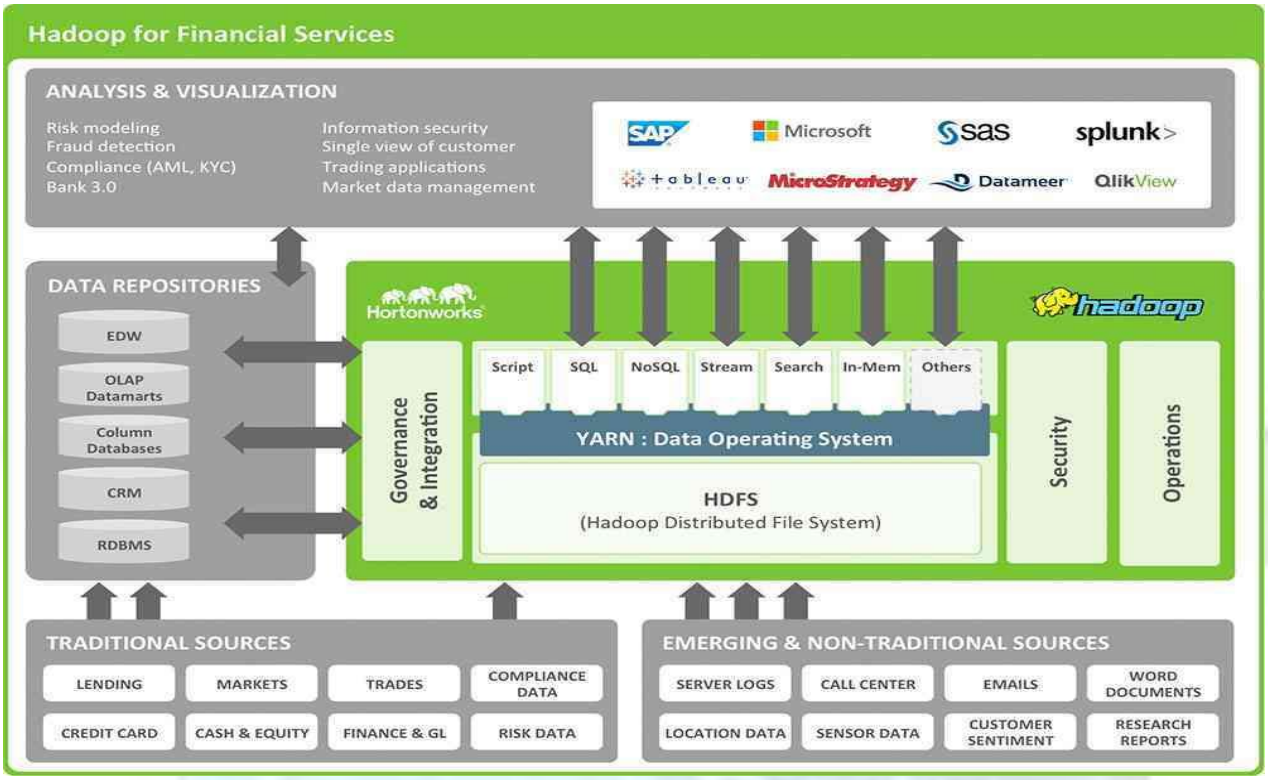

Fig 1. Hadoop for FinancialServices_Diagram

I hold that the catalyst for this disruption is Predictive Analytics – which provides both real time and deeper insight across hundreds of myriad of scenarios –

- Predicting customer behavior in realtime,
- Creating models of customer personas (micro and macro) to track their journey across a Bank's financial product offerings,
- Defining 360 degree views of a customer so as to market to them as one entity,
- Fraud detection
- Risk Data Aggregation (e.g Volcker Rule)
- Compliance etc.

5. CONCLUSION

In this paper we investigate how big Data analytics yields on decision-making support, planning and strategy. The entire financial system can realize benefits from utilizing big data technologies. To successfully identify and implement big data solutions and benefit from the value that big data can bring, financial need to devote time, allocate budget and resources to visioning and planning. With the help of Hadoop the goal of effective citizen care management can be achieved by providing an effective data driven services to citizens by predicting their needs based on the analysis of survey conducted among different classes of citizens.

REFERENCES

1. Yanglin Ren, Monitoring patients via a secure and mobile healthcare system, IEEE Symposium on wireless communication,2011

 DOI:10.22362/ijcert/2016/v3/i12/xxxx

2. Dai Yuefa, Wu Bo, GuYaqiang ,Data Security Model for Cloud Computing, International Workshop on Information Security and Application,2009

3. Jeffrey Dean and Sanjay Ghemawat,MapReduce: Simplified Data Processing on Large Clusters,ACM,2008

4.Kayyali B, Knott D, Van Kuiken S. The big-data revolution in US health care: accelerating value and innovation. McKinsey & Company. 2013 Apr. URL: https://digitalstrategy.nl/wp-content/uploads/E2-2013.04-The-big-data-revolution-in-US-health-care-Accelerating-value-and-innovation.pdf [accessed 2016-11-11] [WebCite Cache]

5. Cong Wang, Privacy-Preserving Public Auditing for Secure Cloud Storage,IEEE,2010

6. Konstantin Shvachko, HairongKuang, Sanjay Radia, Robert Chansler ,The Hadoop Distributed File System,IEEE,2010

7. Bill Hamilton, Big Data Is the Future of Healthcare, Cognizant white paper, 2010.

8. White Paper by SAS, How Government are using the Power of High Performance Analytics, 2013.

9. Michael Cooper &Peter Mell, Tackling Big Data Problems, NSIT Computer Security Workshop

10. Jean Yean, Big Data Bigger Opportunities, White Paper, U.S.General Services Administration.

11. Kruse CS, Goswamy R, Raval YJ, Marawi S Challenges and Opportunities of Big Data in Health Care: A Systematic Review JMIR Med Inform 2016;4(4):e38 URL: https://medinform.jmir.org/2016/4/e38 DOI: 10.2196/medinform.5359, PMID: 27872036 ,PMCID: 5138448.

