A Survey on Big Data in Financial Sector

Anku Jaiswal
Lecturer, Department of Computer and Electronics
Advanced College of Engineering and Management
Lalitpur, Nepal
jaiswalaku@gmail.com

Purrushottam Bagale
Lecturer, Department of Computer and Electronics
Advanced College of Engineering and Management
Lalitpur, Nepal
puru.bagale@gmail.com

***Abstract*— With the advancement of computing technology, there has been rapid increase in volume, variety and velocity of data being produced, which is termed as Big Data. The Big data holds the potential to transform the enterprise operations and processes by leaps and bounds. One of the sectors that can highly benefit from the Big Data is financial. They have access to huge amount of transaction data which can be processed to gain the competitive advantages over their peers, enhance the customer banking experience, risk analysis and mitigation, operation and optimization. This paper discusses about the application of Big Data and its challenges in one of the financial sectors, which is Banking, and the benefit of Big Data in this field.**

Keywords*— *Big Data, Financial Sector, Bankingt, Hadoop, Map Reduce

I. INTRODUCTION

As the name implies Big Data is defined as a voluminous amount of Data which cannot be handled by traditional Database management system. It includes techniques for extracting hidden information and pattern from a large amount of unstructured data [1]. Big data came into the picture with the increase in unstructured data –which does not have a predefined structure or format. The daily statistics of data being generated can be seen live in the site: http://www.internetlivestats.com/ [2]. It shows how fast the data is increasing on hourly or second basis as well as how big data users are increasing daily. Big data users are those who are responsible for generating these data using various technologies. The volume of data being generated has increased to Exabyte, which is very difficult to handle. The data is generated from various sources such as social media sites, phone, daily transactions and financial sectors. Big Data can be defined by the four V'S concept shown in [1, Fig. 1] which includes: Volume, Variety, Velocity and Veracity [3].

Volume: It refers to amount of data being generated and measured in terms of higher number of bytes for example: petabyte or Exabyte. The main source of these data is internet enabled applications such as: e-commerce, social media etc. It is estimated that around 40 zeta bytes of data will be generated by the end of 2020, which is around 300 times of data in 2005.

Variety: It refers to the variety of data generated from different sources. The data generated is not always same as it can be structured, semi structured or unstructured. Big data has come into concept with the increase in unstructured data. It is estimated that 400 million tweets are sent per day and around 30 billion contents are shared daily on Facebook.

Velocity: This includes the analysis of streaming of data, which refers to the speed at which the data is increasing. These days' data is increasing at a rate faster than the processor can process it. Hence here comes the concept of parallel processing. According to a survey "The New York stock exchange captures 1TB of Trade information during each session.

Veracity: It discusses about the quality and uncertainty present in the dataset.

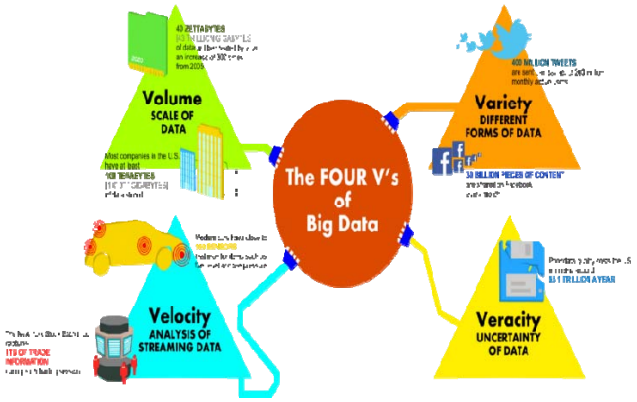

Fig. 1. The Four V's of Big Data

The various sources of the Big Data include [4, Fig. 2]:

Public Data: Data generated from government agencies, weather, research etc.

Transactional Data: Data generated from various transactions from Mobile, Web applications.

Social Media: A huge amount of Data is generated from social Medias such as Facebook, WhatsApp, email etc.

Enterprise Data: Large amount of data in word, excel, pdf, spreadsheet etc format coming from enterprise.

Archives: These are data which are very old or are not required by the organization such as employee's old records, scanned copy etc.

Activity Generated: These are data generated from machine such as images, data from sensors, which are in suppressed form.

978-1-5386-0604-9/17 $31.00 © 2017 IEEE
DOI 10.1109/NaNA.2017.46

Fig. 2. Sources of Big Data

II. WHY BIG DATA

As we can see the tremendous amount of data in every field as shown in [4, Fig.3] concept of big data is a savior with the use of various technologies such as Map Reduce and Hadoop. Also, it can be used with various data mining and Machine learning techniques to extract useful information from the data. Each and every generated data is useful as some important features can be extracted from it. For example, sentiment analysis from tweets, positive and negative comments from newspaper, rating from various ecommerce sites. Efficient use of big data in various sectors helps in proper decision making by various decision makers and also to quickly access risk and take proper action in time. Financial Institute has been gathering large amount of data over a period of time using traditional data approach. The conventional data management techniques are no longer sufficient to handle the massively large, high-velocity, heterogeneous financial data. Big Data analytics and reporting tools, handle extremely large-scale, often real-time, heterogeneous data sets. Hence, it is very important for the financial organizations to use their Big Data investment strategies with focus on their business-specific goals, risk management, product innovation, risk and market intelligence, cost reduction, services, operations, and many more [5].

Fig. 3. Generation of Data

III. TOOLS IN BIG DATA

Various tools and techniques are used in Big Data to handle voluminous amount of data, some of them are:

Map Reduce is a software framework that allows developers to write programs that process massive amounts of unstructured data in parallel across a distributed cluster of processors or stand-alone computers [6]. Map Reduce allows data to be distributed across a large cluster, and can distribute out tasks across the data set to work on pieces of it independently, and in parallel. This allows big data to be processed in relatively little time. Apache has produced an open source Map Reduce platform called Hadoop.

Map a function that parcels out work to different nodes in the distributed cluster. Reduce another function that collates the work and resolves the results into a single value. The Map Reduce framework is fault-tolerant because each node in the cluster is expected to report back periodically with completed work and status updates. If a node remains silent for longer than the expected interval, a master node makes note and re-assigns the work to other nodes as shown in [7, Fig. 4].

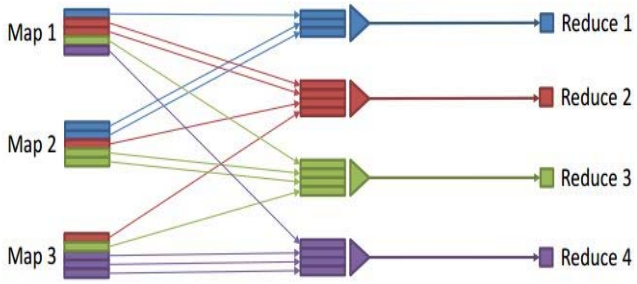

Fig. 4. Map Reduce

Apache Hadoop is a framework that allows the distributed processing of large data sets across cluster of commodity computers using simple programming model. Developed by Apache as an open source distributed platform it is based on Google's Map Reduce and runs on a Java architecture framework that supports the processing of large data sets in a distributed computing environment. Hadoop allows businesses to process large amounts of data quickly by distributing the work across several nodes and is good for big data sets on large cluster [8].

IV. BIG DATA IN FINANCIAL SECTOR

Big data is the term for a collection of data set so large and complex that it becomes difficult to process using on-hand database management tool or traditional processing application. Sectors like retail industries, finance, health care and media generate high volume of data in petabytes due to high churn of transaction and inventory. These companies can utilize big data insight to ensure higher portability. In all these sectors data is one of the most important assets as it helps in decision-making and risk management [9]. Financial sector

such as banks are using Big Data to change their business process, their organization and entire industry. At the end of the day banking sectors end up conducting hundreds of transactions daily as shown in [11, Fig. 5], hence generating a pool of data. An unsolved question of these firms is how to analyses these data and gets useful information for decision-making. Financial sectors are utilizing big data to predict the stock's movement and securities in market by developing predictive algorithm. Percentage of Bank using Big Data is shown in [12, Fig. 6]

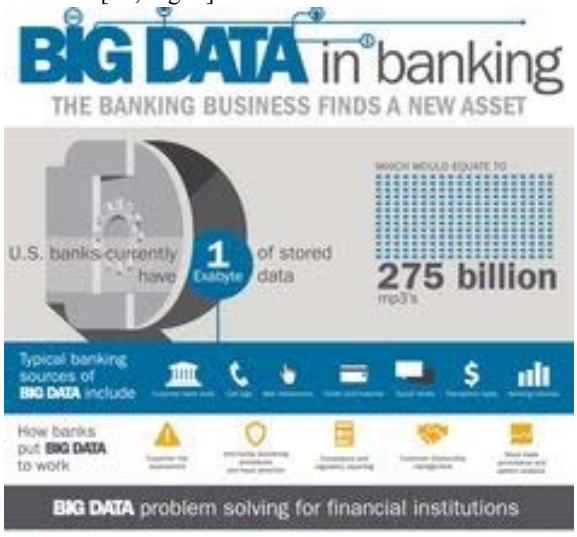

Fig. 5. Big Data in Banking Sector

Another area in which big data is a very useful tool is in analysis of large amount of complex stock market exchange data and using it to make critical financial decision [10]. Intelligent decision system with machine learning and data mining can be used to forecast the stock market. With a big data platform, stock market traders and investment portfolio manager can process vast amount of unstructured data to identify the best compliance in which to invest. The finance companies uses unstructured public information like news, product review, supplier data and price list changes and processes it with big data technology thus producing mathematical model that help traders decide which stock to buy or sell. Yet another important field of using big data is extracting sentiments from news articles, which concludes the article being positive, negative or neutral [11]. Big data scans through all financial transactions taking place in an organization; this ensures the real-time detection of fraudulent activities. Any alteration to normal behavior is instantly caught. Segmentation process contextual marketing which always results in greater returns compare to broad spectrum marketing. Big data enables the classification of customer into various segments based on consumer behavior across multiple touch point including website. According to a survey, around 30% of companies and financial sectors have started implementing big data and targeted on techniques to use big data in various ways to mitigate risk and achieve more benefit from generated data. Some trends to adopt Big Data in financial sectors include larger dataset from various sources including historical data, strategies to improve enterprise transparency and avoid risk, development of predictive model to discover consumer behavioral pattern and to minimize challenges related to data security [12].

Fig. 6. Percentage of Bank using Big Data

V. Big Data Deployment

Big data is really useful concept for the banking sector. The various ways in which banks can use big data technologies in their business processes are as follows:

- To provide real time recommendations to the customers based on their changing financial profile. It can be very useful for the customers with very low income as well as high income to adapt to the policy that meets their current financial requirement.
- Categorize and segment each and every customer for the target marketing and campaigns as per the client requirements. Thus big data can help the banks to provide more personalized services to their customers.
- Fine-tune each and every parameter that can affect the strategies to be implemented for the achievement for financial goals and objectives.
- Can help to control the costs by identifying the unproductive factors and schemes. There are many unproductive schemes and factors that banks may be practicing, which can be easily identified and mitigated using the big data technology.
- Identify the anomalies and outliers in the transactions for the early detection of frauds. As fraud in the transaction is big headache and concern for the banks, big data provides effective remedy in such cases.

Thus, big data provides very good opportunity for the banks to change their business processes and the way they provides the

service to their customer.

VI. Challenges in Big Data

One of the biggest challenges for business intelligence in the banks are the inability to use data for organization decision making as thousands of data is generated on daily basis [13]. This has become even more challenging with the four V'S of big data which include the velocity, veracity, volume and variety which cannot be handled using traditional database [14]. Also for handling of such types of data requires new techniques such as Hadoop, NoSQL, Map Reduce, which is not known to many financial organizations. Some of the limiting factor in this domain includes lack of standard data, lack of transparency and data quality gaps across different financial unit. Hence data has been generated for years but not analyzed due to lack awareness, skilled and qualified big data analyst and risk associated with sensitive financial information.

VII. Conclusion

Data is rapidly increasing due to use of internet and technology; hence it is very important to use techniques and technology to manage these data and extract useful information and patterns from them. Hence Big Data can be very useful technology in such cases to handle such vast amount of large and complex datasets. Financial institutions are using big data in a variety of ways to deliver better business outcomes for their organizations.

References

[1] Samiddha Mukherjee, Ravi Shaw, Big Data Concepts, Applications, Challenges and Future Scope, internationalJournal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 2, February 2016

[2] http://www.internetlivestats.com/

[3] Rama Devi Gunnam, PratyushaGudavalli, ReshmaPothuri,"Survey on Importance and Tools Used: Big Data", International Journal of Advanced Research in Computer Science and Software Engineering

[4] Rama Devi Gunnam, PratyushaGudavalli, ReshmaPothuri,"Survey on Importance and Tools Used: Big Data", International Journal of Advanced Research in Computer Science and Software Engineering

[5] https://www.cs.helsinki.fi/u/jilu/paper/bigdataapplication04.pdf

[6] VidyullathaPellakuri, Dr.D. Rajeswara Rao, Hadoop Mapreduce Framework in Big Data Analytics, International Journal of Computer Trends and Technology (IJCTT) volume 8 number 3 Feb 2014

[7] https://www.cs.colorado.edu/~kena/classes/5448/s11/presentations/hadoop.pdf

[8] https://www.snia.org/sites/default/education/tutorials/2013/fall/BigData/SergeBazhievsky_Introduction_to_Hadoop_MapReduce_v2.pdf

[9] Analytics_The_real_world_use_of_big_data_in_Financial_services_Mai_2013.pdf

[10] Junfei Qiu, Qihui Wu, Guoru Ding, Yuhua Xu ,Shuo Feng, A survey of machine learning for big data processing , EURASIP Journal on Advances in Signal Processing2016

[11] big-data-in-financial-services-wp-2415760.pdf

[12] https://www.cs.colorado.edu/~kena/classes/5448/s11/presentations/hadoop.pdf

[13] https://www.snia.org/sites/default/education/tutorials/2013/fall/BigData/SergeBazhievsky_Introduction_to_Hadoop_MapReduce_v2.pdf

[14] Analytics_The_real_world_use_of_big_data_in_Financial_services_Mai_2013.pdf

[15] Junfei Qiu, Qihui Wu, Guoru Ding, Yuhua Xu ,Shuo Feng, A survey of machine learning for big data processing , EURASIP Journal on Advances in Signal Processing2016

[16] big-data-in-financial-services-wp-2415760.pdf

[17] https://whitepapers.em360tech.com/wp content/files_mf/1427803213insideBIGDATAGuidetoBigDataforFinance.pdf

[18] Sayali Gaikwad ; Pranali Nale ; Ravindra Bachate, Survey on Big Data Analytics for digital world, http://ieeexplore.ieee.org

[19] Challenges and Opportunities with Big Data: A community white paper developed by leading researchers across the United States