Iberoamerican Business Journal

IBJ /JOURNAL

Vol 7 N° 1 | Julio 2023 p. 76 - 97   ISSN:2521-5817 DOI: http://dx.doi.org/10.22451/5817.ibj2023.vol7.1.11077

How European banks are implementing Big Data Analytics: applications, tools and opportunities

Cómo están implantando los bancos europeos el Big Data Analytics: aplicaciones, herramientas y oportunidades

Kuong Feng [1]

[1] Ivey College of Business, Iowa State University, USA (kuong@gmx.com)

(**Kuong Feng**) [1]

Professional description: Kuong Feng is a graduate student in Master of Science in Business Analytics at the Ivey College of Business, Iowa State University, USA. She has previously attended Shanghai Jiao Tong University where she received her Bachelors in Computer Science. She has deep interest in finding insights using big data and thus she plans to pursue PhD in the Statistical Analytics field.

RESUMEN

El objetivo de este trabajo de investigación es analizar cómo los bancos comerciales indios manejan los big data, que se refiere a un conjunto de datos extremadamente grande que requiere análisis, gestión y validación a través de las herramientas tradicionales de gestión de datos.

Propósito: Los bancos son una de las industrias de servicios financieros que manejan una gran cantidad de datos de transacciones, que deben ser gestionados, examinados y utilizados en beneficio tanto del banco como de sus clientes. Este estudio examinará los factores que tienen un mayor impacto en los bancos a la hora de manejar big data y cómo la analítica puede crear valor para el negocio.

Métodos de investigación: Se recopilaron datos secundarios de diversas fuentes, como artículos, revistas y sitios web. El estudio se centra en la gestión de big data, la gestión de riesgos, la detección de fraudes, la segmentación de clientes y el valor

Recibido: 24 de julio del 2023. Aceptado: 25 de Julio del 2023. Publicado: 31 de julio 2023

Newman
Escuela de Posgrado

empresarial de las industrias bancarias. Se ha desarrollado un marco conceptual para destacar los factores que tienen un mayor impacto en la gestión de big data en la industria bancaria.

Resultados: Los resultados indican que la analítica de big data tiene un impacto significativo en el valor de negocio de los bancos, y se han identificado los factores que influyen en el valor de negocio.

Conclusiones: Utilizando big data y adoptando las tecnologías emergentes, las empresas pueden aumentar el valor de su organización.

Palabra clave: BDA, Machin learning, big data analytics; sector bancario; herramientas de Big Data

ABSTRACT

The objective of this research paper is to analyze how Indian commercial banks handle big data, which refers to an extremely large data set that requires analysis, management, and validation through traditional data management tools.

Purpose: Banks are one of the financial services industries that deal with a vast amount of transaction data, which must be managed, scrutinized, and utilized for the benefit of both the bank and its customers. This study will examine the factors that have a greater impact on banks when handling big data and how analytics can create value for the business.

Research methods: Secondary data was collected from various sources such as articles, journals, and websites. The study focuses on big data management, risk management, fraud detection, customer segmentation, and the business value of banking industries. A conceptual framework has been developed to highlight the factors that have a higher impact on big data management in the banking industry.

Findings: The findings indicate that big data analytics has a significant impact on the business value of banks, and the factors influencing business value have been identified.

Conclusion: By utilizing big data and embracing emerging technologies, companies can enhance the worth of their organization.

Keywords: BDA, Machin learning, big data analytics; banking industry; Big Data tools

I. INTRODUCTION

Big Data has numerous applications including weather & temperature prediction, climate prediction, flood assessment, criminal identification, and the financial services business are just a few of the many uses that create data every day from a wide variety of applications, equipment, and geographical research activities. Current scenarios equate big data with fundamental technologies and companies like Google, Facebook, and IBM that mine massive amounts of data for insights (Ji, 2022). In financial services, a new age of transparency has begun. When it comes to customers care, compliance, and other regulatory obligations, banking is no exception to the fast emergence of big data. As the world's population and average lifespan both rise, so too are the stakes for making data-driven choices on how best to offer financial services. BDA, so named for its various attributes like volume and complexity, and its breadth, depth, has great potential for banking investors. Experts and investors in the banking business have started regularly analyzing big data to generate insight, and that work is in very early stage that still requires development to tackle banking delivery challenges and improve quality. There are already well-established application of BDA in various financial services use cases, such as analyzing disease pattern, customers sickness and calculating financial services costs, outcomes to identify the most effective and cost-efficient treatments (Moore, 2003). The study of banking information is known as banking analytics, which is noted as a culmination of medical, hospital and customers related information to analyze in a proper manner and derive insights. Acquiring, storing, and retrieving data for improved outcomes by banking practitioners is the field of financial services informatics. Data in the banking sector is known for being both diverse and complex.

Variation due to the interconnection of many different types of biological data, such as sensor readings, images, gene arrays, laboratory results, free text, and demographic information (Hoyle, 1995). The number of banking data (including doctor's notes, lab findings, and clinical data), the vast majority of which is unstructured and not kept electronically, is fast growing. The digitalization of these massive stockpiles of paper information is now a primary priority. The data size revolutions are making it harder to accomplish this goal (Hulland, 1999). When it comes to big data, the various kinds of Vs are where the majority of the terminology and models put their concentration. Electronic Financial services Records (EHR), machine-generated/sensor data, financial services information exchanges, customers registries, portals, genetic databases, and public records are all examples of the several types of data used in banking applications. Big data in banking comes mostly from public records, which necessitates effective data analytics to address the corresponding banking issues. A study from 2012 estimates that the volume of banking data was close to 550 petabytes, and that by 2020, it would have grown to close to 26,000 petabytes (Akter, 2022). The process of implementing the translation of unprocessed data into usable insights is challenging due to the big-data sources' varied data formats, enormous volume, and associated uncertainties. Due to the complexity involved, financial services feature identification in financial services data and class attribute selection for financial services analytics need advanced and architecture-specific methods and tools.

II. Big Data Analytics within Financial services Informatics

The key factor that sets apart traditional financial services analysis from big-data financial services analytics is the incorporation of computer programming. Previously, banking organizations had to depend on external sectors to carry out large-scale data analyses. Information technology has earned the confidence of many in the banking industry because it consistently produces useful results. The banking sector is currently struggling to handle the massive increase in banking data. However, the industry could potentially take advantage of the expanding field of big data analytics. A significant portion of the data generated by this system is stored in physical copies that need to be digitized later on.

Supporting cutting-edge treatment for customers, enhancing customers outcomes, and reducing banking expenditures are all possible thanks to data (Ghobadi, 2016). Decisions made by doctors, such as whether or not to do a cardiac procedure, may now be predicted using big data analytics that factor in the customer's age, present state, and financial services status. With the current technology, software, and management tools, handling banking data sets can be difficult. Big data can help with this issue since banking data is continuously increasing and reimbursement practices are changing. As a result, the importance of deliberate application and meritbased compensation has grown in the financial services field. More than 150 exabytes of data were generated by banking industry organizations in 2011 (Hutter, 2018) , all of which must be effectively evaluated for the banking system to benefit in any way (ISCA, 2011). EHRs accommodate several formats for archiving financial services-related information. The discipline of bioinformatics, which deals with the massive amounts of data produced by genome sequencing, has also seen a dramatic uptick in recent years (ISCA, 2011). Several methods of analysis exist for deciphering financial services data for clinical application (Isenberg, 2022). The banking informatics community has a significant challenge as they attempt to create techniques for handling large data due to its many sources and formats. There is a great need for a method that can bring together different types of data (Ji, 2022).

There are a variety of possible conceptual frameworks to use to spot discrepancies in huge data sets from several sources. Examples of frameworks used for analyzing banking data include predictive analytics:

In recent years, predictive analysis has become a popular business intelligence strategy, with practical applications that extend beyond the corporate world. Among the many approaches to analyzing big data, text analytics and multimedia analytics stand out as particularly useful (Ji, 2022). One of the crucial subfields is predictive analytics, which employs statistical techniques such as data mining and machine learning to examine past and present data to predict future outcomes. Today, banks are using predictive approaches to identify customers who may be at risk for readmission (Ji, 2022). Financial services professionals may use this information to improve customers care. Understanding and using machine learning is crucial for predictive analysis.

Machine learning and data mining are closely related in financial services. Both methods search for patterns in data, but machine learning uses the data to improve its understanding. By analyzing the information, machine learning can adjust its behavior accordingly. Electronic financial services records (EHRs) are the most common source of big data in financial services. EHRs contain customers information such as financial services history, allergies, symptoms, and test results. These records are accessible to banking professionals through a secure electronic database, allowing for easy updates and avoiding duplication of information.

III. 4 Vs of Big Data in Banking

4 main characteristics (shown in Fig. one) which are related to great data: amount, variety, velocity, and then veracity.

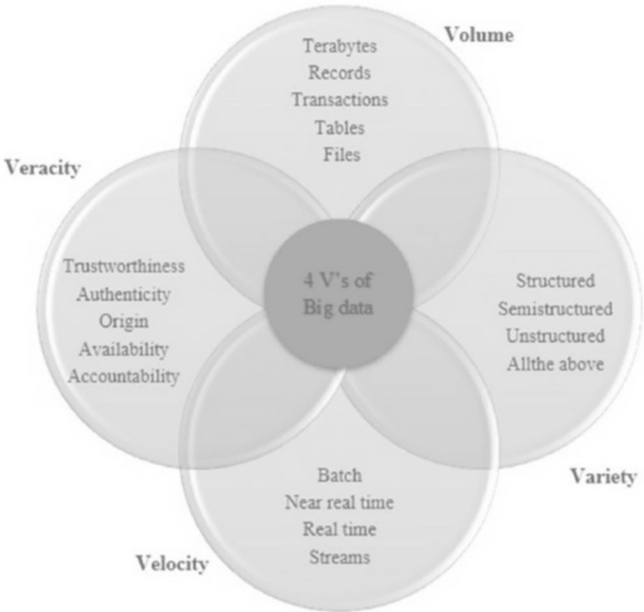
Big data refers to data sets that are very large in volume. The maximum allowable size of such data sets is not known. Massive data sets are those that exceed the capacity of conventional database management systems and information processing architecture (Ji, 2022). The increase in data storage and processing system affordability, alongside the need to gain significant insights from data to improve corporate operations, enhance efficiency, and provide better customer service, has led to a surge in the amount of data generated by modern banking and IT systems.

Velocity: Speed: Speed, the most fundamental data expansion is due to the rate at which new information is being gathered (Ji, 2022). There is a rapid increase in the volume of data produced by banking systems. It is necessary to make a choice based on the processed results of a large and diverse set of structured and unstructured data at a rapid pace.

Variety: Data may be in a wide variety of formats, including text, images, audio, video, and sensor readings. Clinical data (customers record data) is an example of structured data information. The need just that it be gathered, stored, and processed by a single machine. Only around 5–10% of banking data is structured. Emails, images, movies, audio files, and hospital financial services records, doctor's notes, paper prescriptions, and x-ray films are all examples of unstructured or semi-structured data (Ji, 2022).

Data veracity refers to how confident one may be in the data's stated interpretation. The trustworthiness and accuracy of data varies widely depending on the source. Dependable as well as error free major details analytics in banking is important. Unsupervised piece of equipment mastering algorithms, nonetheless, could make judgments according to information which is possibly deceptive or irrelevant. Banking analytics seeks to draw out useful insights through this information to correct clinical results and customers care.

Figure 1: Four dimensions of Big Data



IV. Impact of Big Data within the Banking System

The utilization of big data has the ability to enhance the precision of customers diagnoses and the effectiveness of financial services informatics systems. This will lead to significant impacts on the financial services services infrastructure via five pathways, as illustrated in Figure 2. The aim of this research is to advance customers outcomes along the aforementioned routes.

By "Right Living," we mean that the customers is making progress toward a happier and financial servicesier lifestyle (Ji, 2022). By adopting a lifestyle that prioritizes financial services and well-being, customers may take charge of their conditions and improve their quality of life. Customers have the power to take an active role in achieving a financial servicesy lifestyle through their choices regarding food, exercise, preventive care, and other daily activities. By following this path, customers can receive optimal care as all clinicians work towards the same goals and collect the same information. With the advancement of technology and big data, this approach has become more practical and reliable. By integrating data from various sources, such as financial services devices, public financial services statistics, and socioeconomic information, banking providers can obtain a comprehensive view of the customer's condition. This enables them to conduct targeted research and provide appropriate treatments for their customers.

The Correct Step Forward: Appropriately Recognizing that new diseases, new therapies, and new financial services breakthroughs are ahead (Ji, 2022). Equally significant are developments in the delivery of services to customers.

Improvements in medicine and the effectiveness of research and development will pave the way for innovative approaches to promoting customers financial services and happiness via the nation's universal financial services system (James, 2022). For various reasons, stakeholders need access to trial data as soon as possible. We may utilize this information to investigate promising targets and learn how to enhance standard banking practices.

In order to enhance the quality and value of banking services, banking providers must give diligent and consistent attention to their customers. It is imperative that they achieve the desired outcomes recommended by the customer's financial services plan. To ensure efficient use of data, steps such as identifying and removing data manipulation, waste, and misrepresentation, as well as improving resources, must be taken.

Due to the prevalence of paper records in the banking industry, there is an urgent need for the active digitalization of paper records. Most of this data is also unstructured, making it very difficult for the banking business to derive useful insights The primary focus of the organization is centered around the provision of care to customers, conducting clinical operations, and engaging in research activities. Hadoop ecosystem, a suite of software tools, may assist the banking industry in dealing with this massive data deluge. The banking industry may benefit from the following Hadoop ecosystem applications:

Hadoop Based Applications for Financial services Industry

We know that there are three billion base pairs in human DNA. Fighting cancer requires the effective management of massive data sets. Some cancers are incurable because the mutational patterns that cause them and the body's responses to those mutations differ from person to person. Cancer pattern recognition has led oncologists to the realization that tailoring therapy to each individual's genetic composition is essential. By utilizing Hapdoop's MapReduce technology, a team of researchers at Arizona State University is working on a banking model that can map three billion DNA base pairs to find the most suitable cancer therapy for every individual customers. This model uses a customer's genetic information to determine the most effective course of treatment for their specific type of cancer. This approach offers a framework for therapy that utilizes big data analysis to boost the likelihood of a customer's survival.

Globally, clinics depend on the Hadoop Distributed File System (HDFS) as well as its many parts including Impala, Spark, HBase, as well as Flume to handle the huge quantities of unstructured details produced by affected person financial services methods. The info collected consists of pulse rate, vital signs, respiratory rate, blood sugar level and blood pressure. This particular information wouldn't be available to analysts with no Hadoop. There're around 6,200 pediatric intense hygiene devices (ICUs) in Atlanta, in which kids are able to stay for lengthy times of your time. The kids' important symptoms are continually administered by receptors to ensure which inside the function of an urgent situation financial services related personnel are instantly notified towards the kid's reliability.

To better handle their vast volumes of real-time data on customers care, finances, and payroll, many banks have turned to the NoSQL database available as part of the Hadoop ecosystem. This has allowed them to better pinpoint customers at risk and cut down on routine costs.

Banking intelligence applications utilized by banking providers and payers are supported by Hadoop technology. The Hadoop ecosystem's Pig, Hive, and MapReduce are responsible for processing extensive datasets related to illnesses, medications, symptoms, geographies, opinions, and other factors to extract valuable information, such as the target age, for insurance providers.

Fraud Detection and Prevention: Early adopters of big data analytics in the banking industry are using a variety of strategies to detect fraudulent behavior and develop safeguards against the occurrence of financial services fraud. Hadoop enables businesses to utilize prediction-model applications to detect fraudsters using information such as a person's financial services claims, phone records, earnings, and demographics. Hadoop's NoSQL database can help prevent financial services claims fraud by utilizing real-time financial services apps, legitimate financial services claims, weather forecasting data, voice data recordings, and other relevant data sources. By detecting fraud at an early stage, this technology can save organizations significant amounts of money.

Big Data Analytics Architecture for Financial services Informatics

We know that there are three billion base pairs in human DNA. Fighting cancer requires the effective management of massive data sets. Some cancers are incurable because the mutational patterns that cause them and the body's responses to those mutations differ from person to person. Cancer pattern recognition has led oncologists to the realization that tailoring therapy to each individual's genetic composition is essential. Hapdoop's MapReduce technology makes it possible to map three billion DNA base pairs to find the most effective cancer therapy for each individual customers. A research team at Arizona State University is developing a banking model that uses a customer’s genetic information to determine the best course of therapy for their particular kind of cancer. This approach offers a framework for therapy that utilizes big data analysis to boost the likelihood of a customer’s survival.

The current emphasis in big data analytics is on rather of collecting data, the goal is to learn as much as possible from it and use it (Vinzi, 2022). Extracting useful knowledge, patterns, and information from large, complicated data sets is the goal of data analytics, which entails the creation and use of algorithms for doing so. Since scientists function identifying the easiest way to carry out huge details analytics within financial services methods, a four layer structure, incorporating a transformation level, information supply level, large details wedge level as well as analytic level, was under thing to consider for a few moment. Through this tiered phone system, various details sorts are saved in various methods. Each and every level on the HDFS has a special details processing abilities, utilizing the MapReduce processing strategy. Another levels produce accounts, pass queries along, carry out information mining and also handle online analytics.

The primary need in analyzing big data processing is to bundle data quickly, cutting down on processing time as much as possible. Big data analytics' next big thing is real-time query updates and transformations, or so the thinking goes (Falcioni, 2022). The 3rd demand within serious details analytics is employing & efficiently handling storage space room. The final major details analytics have to have would be to rapidly perfect the ever changing work notations. When comparing big data processing, typical banking information systems and big data analytics frameworks have notable differences (Wanga, 2022). The existing banking system relies on conventional tools placed on a single, isolated machine, such as a desktop computer, for the processing of data. However, large data uses clustering for processing, and it does so by probing the network at numerous nodes. Parallel processing of big quantities of banking details enables the strategy to succeed (twenty four). The financial services business could use open source frameworks as HBase Avro, Sqoop, Pig, MapReduce, Hadoop and Hive to control overall financial services information collections.

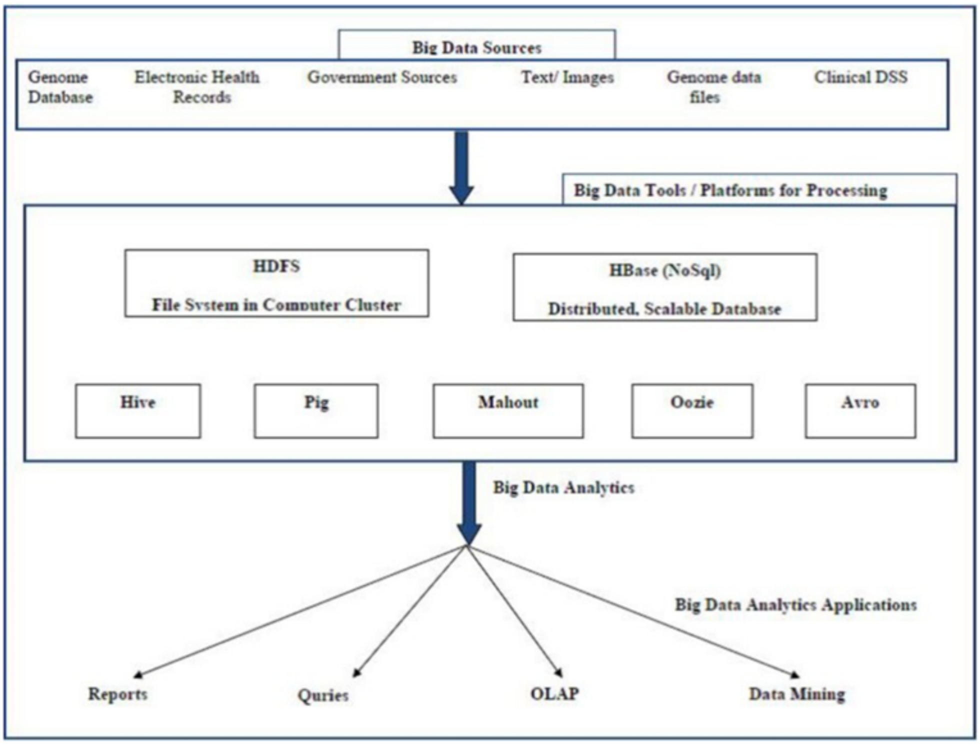
The term "big data technology" encompasses a wide range of development that mirrors the techniques employed with massive datasets (Joreskog, 2022). The first part of this puzzle is the need for massive data storage and processing facilities. The second part is where high performance clusters with a centralized big-data processing architecture really shine (Johnsen, 2022). Using the MapReduce paradigm, the most popular solutions for processing big data analytics provide the safety, scalability, and management of the data. Within the Storage domain name for combining directories accessible via different apps, the 3rd element of serious details analytics apps is (twenty six). The 4th part has essentially the most widely used large details analytics uses within financial services methods, like accounts, OLAP, queries as well as information mining.

Find out, for instance, Fig. There're numerous locations that individuals are able to obtain info regarding individuals, such as EHRs, genomic directories, genome information documents, images and text (unstructured details sources), financial services choice assistance methods, federal organizations, banking examination labs and also drug stores, as well as wellness insurance companies. These specifics are usually accessible within a selection of program tables created within ASCII or textual content structure and then stored in several unique locations. Listed here are a number of the big data Hadoop processing aids which are accustomed for making financial services related uses.

Hadoop's Techniques and Tools for Big Data

Big data that does not conform to any preexisting database structure requires unique tools for management. The IT industry use the Hadoop platform for the many different techniques designed to capture, structure, and analyze such massive datasets (Rao, 2022). To get useful insights from massive data, we need better methods. MapReduce, Mahout, and Hive are only few of the technologies that are part of the Apache Hadoop architecture (Porter, 1985). Below, we'll talk about the many methods for handling large banking datasets and the technologies that make it possible.

Figure 2: Big Data sources and techniques



Apache Hadoop: the word Hadoop has come to imply a variety of things over the years (Fayyad, 2022). It began life as a standalone program in 2002, designed to power a search engine. Since then, it has evolved into a suite of programs for analyzing many kinds of big data (Graen, 1999). Hadoop has evolved from a single project into a methodology for handling data in a way that is quite different from the conventional relational database approach (Fayyad, 2022). Hadoop is an open-source framework and ecosystem that enables the collection, storage, and analysis of massive amounts of data, including but not limited to: Internet images, audios, videos, and sensor records, in both structured and unstructured forms (Wanga, 2022). HDFS facilitates data streaming, the practice of reading huge datasets in one go from a storage medium (Joreskog, 2022). HDFS uses 64 MB or 128 MB blocks. Name nodes and data nodes are the two primary kinds of nodes. To store and retrieve information from the data nodes, a single name node is responsible for all the necessary metadata (Ji, 2022). The name node does not really hold any information. There is a consistent size and order to the blocks used to store files (Johnsen, 2022). HDFS's strong points are its dependability and decentralized design. Information management and file archiving are displays Hadoop's hardware architecture, which includes MapReduce, HBase, and

HDFS. Big data processing is HDFS's bread and butter (Falcioni, 2022). Though it is able to help support several owners simultaneously, HDFS isn't a genuine parallel file program. Instead, a lot of the concurrency as well as coherency overhead demands associated with a genuine parallel file device are calm since the look assumes a tremendous file write once or read many paradigms.

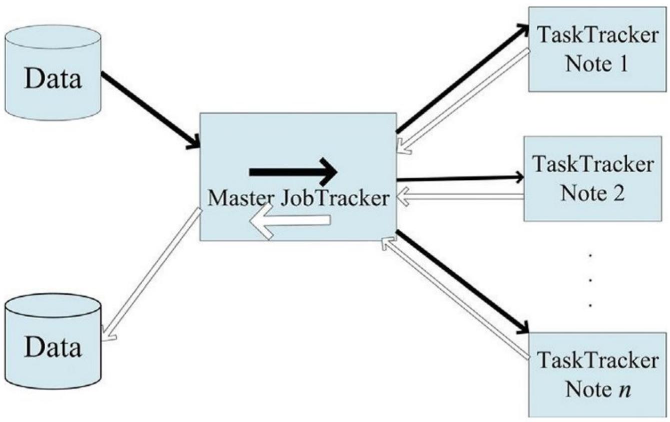
Computing using MapReduce is often connected with Apache Hadoop. MapReduce, a kind of computational model, is more widely employed in the financial services industry than most people think. The idea behind it is straightforward (Joreskog, 2022). Both a "mapping" stage and a "reduction" step exist in MapReduce. In the mapping phase, the input data undergoes a mapping operation. When counting, we enter the reduction phase. Each and every phase on the MapReduce programming paradigm covers key value pairs like output (Isenberg, 2022) : and input A mapping phase which gets feedback inside the kind of crucial benefit pairs and also creates result within the type of key value pairs along with another minimization phase. In Hadoop, the term for the process that divides data into segments of a predetermined size is input splits (Vinzi, 2022). The mapper's value pairs and key are both produced by the Map function. We combine keys that are same. In Fig. 1 we see a simple version of MapReduce in action

Apache Hive is a data warehousing layer on top of Hadoop that allows for SQL-like procedural language analysis and querying (Gefen, 2000). Querying, summarizing, and analyzing data on the fly are all possible with Apache Hive. For SQL-based searches on Hadoop's petabytes of data, Hive is the de facto standard because of its streamlined data extraction, transformation, and access to Hadoop's distributed file system (HDFS) and HBase (Flavian, 2006). To analyze huge data more effectively, many people are turning to open-source tools like Apache Pig. Instead of using MapReduce, you may use Pig (Fornell, 1981). Pig, originally designed as a research project by the Yahoo online service provider, lets users create their own user-defined functions and supports many common data operations.

HBase is a column-oriented NoSQL database in Hadoop (Manyika, 2011) that allows users to store massive amounts of data in both rows and columns. HBase allows for ad hoc read/write capability. Unlike HDFS (Manyika, 2011) , it allows for changes at the record level. HBase uses underlying distributed file systems on commodity machines to store data in a parallel fashion. The preferred file system is frequently HDFS since HBase and HDFS are tightly integrated (Manyika, 2011). HBase is the best option to take use of a structured, low-latency view of Hadoop's massive data stores. It can handle petabytes of data on thousands of nodes with its open-source code scaling linearly.

A complicated system, a tightly designed system, or a network of stations with mutual data dependencies all need the use of an advanced technology known as Apache Oozie. Apache Oozie can manage and execute several Hadoop-related tasks. Workflow engines store and run workflow collections of Hadoop-based tasks, while a coordinator engine handles workflow jobs in accordance with the process schedule. With Oozie, you can build and manage your Hadoop tasks as a workflow, where the results of one operation become the input for the next (Keim, 2000). Oozie is not a suitable replacement for the Yarn planner. Directed acyclic graphs (DAGs) of activities are how Oozie workflow tasks are displayed (Santhanam, 2003). Clients submit tasks to Oozie, which acts as a service in the cluster, and the jobs are either proactively or reactively executed.

Figure 3: How job tracker works



It enables communication between applications created in different languages (Schulte, 2000). It's a common way to link Flume data flows together. Due to its schema-based design, the Avro system decouples the reading and writing processes from the underlying programming language. If your data already has a schema, Avro can serialize it for you (Manyika, 2011). It's a system for transferring data between Hadoop nodes and for making RPCs between client applications and Hadoop services.

Apache Zookeeper is a centralized system that applications may utilize to manage and organize a banking infrastructure and in the spaces between them (Zhang, 2017). It stores the configuration data and the hierarchical name space, both of which are essential in large cluster setups. Various programs may make use of these services to coordinate Hadoop clusters' distributed processing. Reliability in applications is another area where Zookeeper excels (Zhong, 2016). When an application master fails, zookeeper creates a replacement to continue running applications. Hadoop Yarn, built on top of Apache Yarn, is an example of a non-MapReduce application for Hadoop (Zhu, 2019). Yarn is a distributed shell application. The Resource Manager (RM) of Yarn coordinates the allocation of all cluster resources necessary for tasks, while the Node Manager (NM) of Yarn monitors the resources available on each node in the cluster. Job scheduling and container, memory, CPU, and I/O management are both handled by these two parts.

Apache Sqoop: Apache Sqoop is a great application which extracts information coming from the Relational Database Management System (RDMS) and also presents it within the Hadoop structure for query processing. This particular treatment make use of the MapReduce paradigm or any other industry standard solutions as Hive (Davis, 2011) to accomplish this. Hadoop plans could after that use the info kept in HDFS. When it comes to gathering data and transferring it from various computers to HDFS (Donthu, 1998) , you may rely on Apache Flume. Several flume agents may be required to move data across a network of computers and storage facilities. Log files, social networking data, and email are commonplace inputs for Flume.

V. Contributions of the Study

The study evaluates the impact of big data analytics capabilities (BDAC) on financial services industry and its various dimensions. By doing so, it provides a comprehensive understanding of the structure of sustainable banking and BDAC, which can guide further research in other related fields such as marketing and customer behavior. The authors also identify major challenges and barriers faced by the European banking, which can inspire researchers to explore similar issues in other industries and countries. Lastly, the use of the UN online-self assessment and relevant literature provides academic support for the study's methodology.

VI. Implications and recommendations

The paper's findings on the benefits of using big data analytics (BDA) for the sustainability of the banking in Europe could be of great interest to UN global compact specialists, as many European banking companies participated in the online self-assessment that was recommended by the UN. However, it should be noted that many European banking companies were not aware of this assessment before participating in the research. The paper emphasizes the importance of banking companies investing in BDA and creating a well-organized system for it. The study sheds light on the current state of BDA capabilities and sustainability of the banking in Iran, which could be helpful to managers and consultants who are involved in sustainable development. The paper also suggests that companies should establish a secure information sharing system to facilitate an integrated banking.

The potential for effective management of big data analytics (BDA) exists, but it is currently disorganized and not well-coordinated with employee expertise and company policies, resulting in wastage of resources. Managers must seek solutions by examining the microstructure of BDA planning, investment, coordination, and control to ensure efficient BDA management. However, the research model has limitations, including its cross-sectional nature, context specificity, and the need for objective measures to investigate actual impact. Testing the model on other banking and countries would enhance the understanding of the relationship between BDAC and sustainability of the banking.

This paper only examines certain factors that impact a company's strategies, decision-making, and sustainability within the banking. Other factors, such as the company's location, position in the banking, size, partnership status, type of products, organizational culture, level of innovation, profitability, and market share, could be explored in future research. Additionally, further investigation into why certain aspects of the study model were removed is suggested.

VII. Conclusion

We've discussed in information how large details has an effect on the financial services program and also the 4 Vs of great details, & we've additionally provided an introduction to big details within the banking process within this post. We've additionally recommended a conceptual structure for fixing financial services issues in great information making use of Hadoop based terminologies. the entails making use of large details produced by different amounts of financial services related details as well as producing methods for examining this particular information to access answers to financial services concerns. Big data analytics applied to the banking industry. With the use of analytics, doctors can tailor treatment plans to the needs of each customers, giving them better outcomes than would be possible with a one-size-fits-all approach. Big data analytics is still in its infancy, and existing tools and approaches do not enough to address the challenges it presents. As with any large system, big data poses formidable difficulties. Therefore, a lot of study in this area is necessary to find answers to the problems plaguing the banking system.

VIII. References

[1] F. Ghobadi, M. Rohani, “Cost Sensitive Modeling of Credit Card Fraud using Neural Network strategy”, 2016 Signal Processing and Intelligent Systems (ICSPIS), International Conference of pp. 1-5. IEEE.

[2] Hesterberg, T., Moore, D.S., Monaghan, S., Clipson, A. and Epstein, R. (2003), “Bootstrap methods and permutation tests”, in Moore, D.S. and McCabe, G.P. (Eds.) Introduction to the Practice of Statistics (5th edition), available at: http://bcs.whfreeman.com/pbs/cat_160/PBS18.pd f

[3] Hoyle, R.H. (Ed.), (1995) Structural Equation Modeling, SAGE Publications, Inc., Thousand Oaks, CA.

[4] Hulland, J. (1999), “Use of partial least squares (PLS). Strategic management research: a review of four recent studies”, Strategic Management Journal, Vol. 20, No. 4, pp.195–204.

[5] Hutter, T., Haeussler, S. & Missbauer, H., 2018. Successful implementation of an order release mechanism based on workload control: a case study of a make-to-stock manufacturer. International Journal of Production Research , 56(4), pp. 1565-1580.

[6] IBM Tech Trends Report (2011), available at: http://www/ibm.com/developerworks/techtrends report

[7] ISACA (2011), “Data Analysis – A Practical Approach”, available at: http://www.algaonline.org/DocumentCenter/View /2847

[8] James, L. (2010), “Data Analysis for the Retail Industry”, 24 Oct, available at: http://www.yellowfinbi.com/YFCommunityNews Data-Analysis-for-the-Retail-Industry-Part-1 100068

[9] Esposito Vinzi, V., Trinchera, L. and Amato, S. (2010), “PLS path modelling: from foundations to recent developments and open issues for model assessment and improvement”, in Esposito Vinzi, V., Chin, W.W., Henseler, J. and Wang, H. (Eds.), Handbook of Partial Least Squares: Concepts, Methods and Applications, Springer Berlin Heidelberg, Berlin, pp.47–82.

[10] Falcioni, J.G. (2013), “Smart factories, smarter engineers”, Marketing Engineering, available at: http://memagazineblog.org/2013/10/07/smarter factories-smarter-engineers/

[11] Fayyad, U.M., Piatetsky-Shapiro, G. and Smyth, P. (1996), “From data mining to knowledge discovery: an overview”, in: Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P. and Uthurusamy, R. (Eds.), Advances in Knowledge Discovery and Data Mining, MIT Press, Massachusetts, pp. 2-18.

[12] Ji, W. & Wanga, L., 2017. Big data analytics based fault prediction for shop floor scheduling. Journal of Manufacturing Systems, Volume 43, pp. 187-194.

[13] Johnsen, M. (2013), “Using big data to drive retail personalization”, Drugstorenews.com, available at: http://www.drugstorenews.com/article/using-big data-drive-retail-personalization

[14] Jöreskog, K. and Sorbom, D. (1993) LISREL. VIII Scientific Software, Chicago, IL.

[15] Keim, D. A., Mansmann, F., Schneidewind, J. and Ziegler, H. (2006), “Challenges in Visual Data Analysis”, IEEE Information Visualization.

[16] Keim, D.A. (2009), “Visualization techniques for mining large databases: a comparison”, IEEE Transactions on Knowledge and Data Engineering, Vol. 8 No. 6, pp. 923–937

[17] Lohmöller, J.B. (1989), “Latent Variable Path Modeling with Partial Least Squares”, Physica Verlag, Heidelberg. 50. Lohr, S. (2012), “The Age of Big Data”, The New York Times, available at: http://www.nytimes.com/2012/02/12/sunday review/big-datas-impact-in-the-world.html?

[18] 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 Report, available at: http://www.mckinsey.com/insights/business_tech nology/big_data_the_next_frontier_for_innovation

[19] Pinsonneault, a. & kraemer, k. L. 1993. Survey research methodology in management information systems: an assessment. Journal of management information systems, 75-105.

[20] Flavian, C. and Guinaliu, M. (2006), “Consumer trust, perceived security and privacy policy: Three basic elements of loyalty to a web site”, Industrial Management & Data Systems, Vol. 106 No. 5, pp. 601-620.

[21] Fornell, C. and Larcker, D. (1981), “Evaluating structural equitation models with unobservable variables and measurement errors”, Journal of Marketing Research, Vol. 18 No. 1, pp. 39–50.

[22] Gefen, D., Straub, D. and Boudreau, M. (2000), “Structural equation modelling techniques and regression: guidelines for research practice”, Communications of the Association for Information Systems, Vol. 7, No. 7, pp.1–78.

[23] Glaser, B.G. and Strauss, A.L. (1967), The Discovery of Grounded Theory: Strategies for Qualitative Research, Transaction Publishers, Chicago.

[24] Graen, M. (1999), "Technology in Manufacturer/Retailer Integration between Wal Mart and Procter & Gamble, available at: http://citebm.business.illinois.edu/IT_cases/PG Graen.htm 35. Haenlein, M. and Kaplan, A.M. (2004), "A beginner's guide to partial least squares analysis", Understanding Statistics, Vol. 3 No. 4, pp. 283 297.

[25] Porter, m. E. & millar, v. E. 1985. How information gives you competitive advantage. Harvard business review, reprint service.

[26] Ramaswamy, V. and Gouillart, F.J. (2010) The Power of Co- creation: Build it with them to Boost Growth, Productivity, and Profits, Simon and Schuster, Noida.

[27] Rao, A. M., & rothstein, m. A. How analytics is driving the banking innovation in north america., business & IT, 2022

[28] Santhanam, r. & hartono, e. 2003. Issues in linking information technology capability to firm performance. Mis quarterly, 27, 125-153.

[29] SCHULTE, R. (2000), "Application integration scenario: how the war is being won", in Gartner Group (Ed.), Application Integration – Making E-Business Work, Gartner Group, London.

[30] Zhang Y, Ren S, Liu Y, Si S. A big data analytics architecture for cleaner manufacturing and maintenance processes of complex products. J Clean Prod 2017;142. https://doi.org/10.1016/j.jclepro.2016.07.123.

[31] Zhang Z, et al. Pathologist-level interpretable whole-slide cancer diagnosis with deep learning. Nat Mach Intell 2019;1(May). https://doi.org/10.1038/s42256- 019-0052-1.

[32] Zhong RY, Newman ST, Huang GQ, Lan S. Big Data for banking management in the service and manufacturing sectors: challenges, opportunities, and future perspectives. Comput Ind Eng 2016;101. https://doi.org/10.1016/j. cie.2016.07.013.

[33] Zhu K, Li G, Zhang Y. Big data oriented smart tool condition monitoring system. IEEE Trans Ind Inform 2019;16(6):1. https://doi.org/10.1109/tii.2019.2957107.

[34] Davis, D.F., Golicic, S.L. and Boerstler, C.N. (2011), "Benefits and challenges of conducting multiple methods research in marketing", Journal of the Academy of Marketing Science, Vol. 39 No. 3, pp. 467-479.

[35] Donthu, N. and Yoo, B. (1998), "Retail productivity assessment using data envelopment analysis", Journal of Retailing, Vol. 71 No. 1, pp. 89–105.