Big Data Management in Commercial Banking: A Developmental Appraisal of Relevant Underpinning Theories

EFUNTADE, Olubunmi Omotayo, Ph.D.
Department of Economics, Federal University Oye-Ekiti, Ekiti State, Nigeria.
Email: olubunmi.efuntade@fuoye.edu.ng ORCID: 0000 0001 9935-026X

EFUNTADE, Alani Olusegun, Ph.D.
Federal University Oye-Ekiti, Ekiti State, Nigeria.
Email: alani.efuntade@fuoye.edu.ng
ORCID: 0009 0006 6945 4061
DOI: 10.56201/ijbfr.v10.no7.2024.pg61.74

Abstract
This is a conceptual and theoretical review paper that explore big data which is referred to large volumes of structured, semi-structured, and unstructured data that inundates a business on a day-to-day basis and management in banking sector of Nigeria. Big data draws heavily from information theory, which quantifies the amount of uncertainty involved in the value of a random variable. Big data analytics helps banks identify and mitigate various risks, including credit, market, operational, and regulatory risks. By analyzing historical data and real-time transactions, banks can detect fraud, predict default rates, and assess the creditworthiness of borrowers more accurately. Machine learning and data mining propose that by applying algorithms and statistical techniques to big data, banks can uncover patterns, trends, and relationships that can inform decision-making processes, improve risk management, and enhance customer experience. Statistical methods and probability distributions are used to analyze and make sense of large datasets, enabling data scientists to draw meaningful conclusions from the data. Statistical and probability theory propose that by applying mathematical models and methods to large datasets, banks can gain insights into customer behavior, market trends, and financial risks. It is recommended that banks should establish a robust data governance framework and upgrade data infrastructure to harness the power of big data to drive innovation, improve decision-making, and enhance competitiveness in the digital era.

Keywords: *Banking sector, data mining theory, information theory, Nigeria, statistical theory, SWOT analysis*

1.0 Background to the Study

The concept of Big Data gained prominence in the early 2000s, with proponents such as Doug Cutting and Mike Cafarella, who developed the Hadoop framework for distributed processing of large datasets. Other key figures include Google's Jeff Dean and Sanjay Ghemawat, who pioneered MapReduce, a programming model for processing big data (Ahamad, *et al.,*2019).

Internationally, current trends in Big Data include: AI and Machine Learning Integration: Leveraging AI and machine learning algorithms to extract insights from big data. Edge Computing: Processing data closer to its source to reduce latency and bandwidth usage. Privacy and Security: Addressing concerns related to data privacy, security, and ethical use. Real-Time Analytics: Analyzing data in real-time to enable quicker decision-making. Industry-Specific Applications: Tailoring Big Data solutions to specific industries, such as healthcare, finance, and retail (Mhlanga, 2022).

In the developing world, Big Data trends may focus more on leveraging data to address socio-economic challenges, improve healthcare delivery, enhance agriculture productivity, and support sustainable development initiatives. Additionally, there may be efforts to bridge the digital divide and build local capacity in data analytics and management (Mhlanga, 2020).

2.0 Conceptual review

Big data refers to vast volumes of structured, semi-structured, and unstructured data that cannot be processed or analyzed using traditional methods. This data is characterized by its volume, velocity, and variety (Ahamad, *et al.,*2019).

Big data refers to large volumes of structured, semi-structured, and unstructured data that inundates a business on a day-to-day basis. This data comes from various sources such as social media, sensors, devices, and other digital sources. The three key characteristics of big data are volume, velocity, and variety, often referred to as the 3Vs (Mhlanga, 2021). The goal of big data analysis is to uncover hidden patterns, correlations, and other insights to inform decision-making processes and gain a competitive advantage.

Big data management involves the collection, storage, processing, and analysis of large and complex datasets to extract valuable insights and support decision-making processes. It encompasses various techniques, tools, and technologies to handle the challenges posed by big data (Sivarajah *et al*., 2017).

The components of big data concepts include: Data Sources: Where the data originates from. Data Storage: Systems and technologies used to store large volumes of data. Data Processing: Techniques and tools for processing and transforming raw data into usable formats. Data Analysis: Methods for extracting insights and patterns from data. Data Visualization: Presenting data and insights in a visually understandable format (Ahamad, *et al.,*2019).

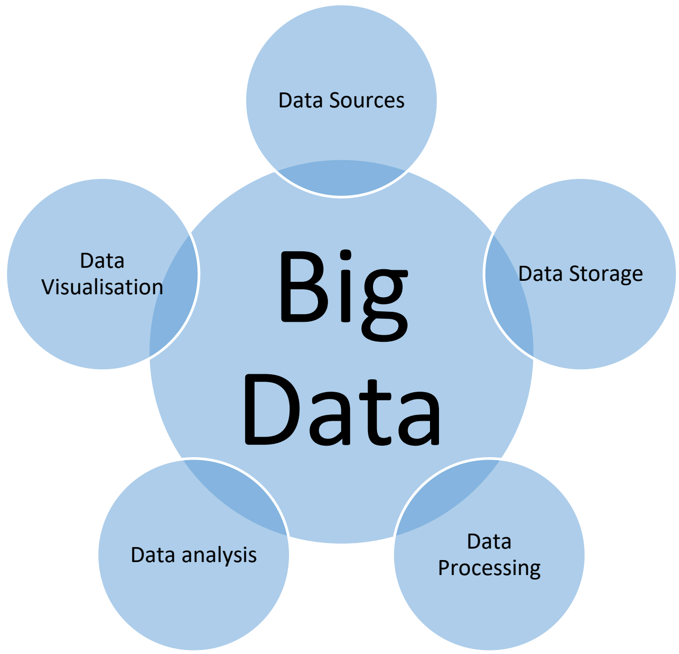

Fig. 1: Basic steps in Big Data Management
Source: Authors' Conceptualisation, 2024

Schematic steps in Big Data Management typically, as depicted in Fig. 1 include data acquisition, data storage, data processing, data analysis, and data visualization. These steps form a cyclical process where insights gained from analysis inform subsequent data acquisition and processing. These steps are often iterative, as organizations continuously collect, store, process, analyze, and utilize big data to drive innovation and improve operations. developing economies, Big Data management in the banking sector involves leveraging large volumes of data to gain insights into customer behavior, improve risk management, enhance operational efficiency, and develop personalized financial products and services. This typically includes collecting, storing, analyzing, and interpreting data from various sources such as transaction records, customer demographics, social media, and digital interactions. Implementing effective Big Data management strategies can help banks in developing economies better understand their customers, make data-driven decisions, and stay competitive in an evolving financial landscape. Additionally, it can also facilitate financial inclusion by reaching underserved populations and tailoring services to their specific needs.

Fig 2: Areas where Big Data management is relevant in banking transactions and systems
Source: Authors' Compilation, 2024

The conceptual view of Big Data emphasizes the importance of harnessing large datasets to gain actionable insights, drive innovation, and improve decision-making processes across various domains, including business, healthcare, finance, and government.

3.0 Theoretical review and underlying theories of big data management in banking sector

Advocates of big data argue that by harnessing the power of large datasets, organizations can gain valuable insights that can drive innovation, improve efficiency, and enhance decision-making processes (Ahmad *et al.,* 2018). Assumptions underlying big data analysis include the belief that more data leads to better insights, that correlations discovered in the data are meaningful, and that algorithms can uncover valuable patterns without prior hypotheses. Propositions of big data include the idea that data-driven decision-making can lead to competitive advantages, that predictive analytics can anticipate future trends and behaviors, and that real-time analysis of streaming data can enable proactive responses to events as they unfold (Ahamad, *et al.*,2019).

The theoretical and conceptual frameworks of big data have evolved over time, with contributions from various disciplines including computer science, statistics, mathematics, and information theory. Key proponents and researchers in the field have continually refined and expanded these frameworks to address the complexities and challenges of analyzing and making sense of large and diverse datasets.

3.1 Information theory

Information theory plays a crucial role in the management of big data by providing a framework for understanding the fundamental principles of data representation, transmission, and processing (Oussous *et al.*, 2018).

Information theory proposes that the value of data lies in its ability to reduce uncertainty. In the context of big data management, this means that the more information we have, the better equipped we are to make informed decisions and predictions. Concepts such as entropy, mutual information, and compression algorithms are central to information theory and are used in various aspects of big data management, including data storage, transmission, and analysis (Zhou *et al*., 2017). Information theory provides a quantitative measure of information content: Information theory allows us to quantify the amount of information present in a dataset using metrics such as entropy, which measures the uncertainty or randomness of the data.

Information theory offers insights into data compression: By leveraging compression algorithms based on information theory, big data systems can efficiently store and transmit large volumes of data while minimizing storage and bandwidth requirements. The theory forms the basis for data encoding and transmission: Information theory principles are used in encoding schemes such as Huffman coding and in communication protocols to ensure reliable transmission of data over networks (Ahmad *et al.,* 2018).

The foundation of information theory was laid by Claude Shannon in his seminal paper "A Mathematical Theory of Communication," published in 1948 (Ahmad *et al.*, 2018). The theory assumes that data can be effectively represented as discrete symbols or bits, allowing for mathematical analysis using information-theoretic measures. Information theory also assumes that reducing uncertainty in the data leads to better decision-making and predictive accuracy. Data sources are independent and identically distributed, although this may not always hold true in practice for big data applications (Oussous *et al.*, 2018).

Information theory often relies on simplified models of data and communication systems, which may not capture the full complexity of real-world scenarios (Sivarajah *et al*., 2017). The theory ignores semantic content: Information theory focuses on the statistical properties of data but may overlook the semantic meaning or context of the information, which can be crucial for certain types of analysis. It assumes perfect knowledge: Information theory assumes perfect knowledge of the underlying data distribution, which may not always be attainable in practice, especially in dynamic and rapidly evolving big data environments (Ahmad *et al*., 2018).

Overall, information theory provides a powerful framework for understanding the principles of big data management, but its application requires careful consideration of its assumptions and limitations in real-world scenarios.

3.2 Statistical and probability theory

Statistical and probability theory are essential components of big data management in banking, enabling financial institutions to analyze vast amounts of data to make informed decisions, assess risks, and detect fraudulent activities. Here's an explanation of statistical and probability theory in the context of big data management in banking:

Statistical and probability theory propose that by applying mathematical models and methods to large datasets, banks can gain insights into customer behavior, market trends, and financial risks (Oussous *et al.*, 2018). This theory offer a framework for analyzing data distributions, estimating probabilities of future events, and making statistical inferences based on observed data.

Statistical and probability models allow banks to quantify and manage various types of risks, including credit risk, market risk, and operational risk. By analyzing transaction data and customer behavior patterns, banks can use statistical methods to detect anomalous activities indicative of fraudulent behavior. Probability theory provides a foundation for decision-making under uncertainty, allowing banks to assess the likelihood of different outcomes and make optimal choices based on probabilistic forecasts (Alzakholi *et al.*, 2020).

Statistical and probability theory assumes that data follows certain statistical distributions, such as normal distribution for risk modeling or Poisson distribution for event occurrence (Sivarajah *et al*., 2017). It assumes that observed data is representative of the underlying population and that statistical inferences drawn from the data are reliable. It further assumes that events are independent and identically distributed, although this may not always hold true for financial data due to autocorrelation and temporal dependencies (Zhou *et al*., 2017).

Statistical models are simplifications of complex real-world phenomena and may not fully capture the underlying dynamics of financial markets or customer behavior. Statistical analysis relies on high-quality, accurate data, but banking datasets may suffer from data quality issues such as missing values, errors, or inconsistencies. Statistical models are sensitive to the assumptions made about the data distribution and model parameters, and deviations from these assumptions can lead to biased results or inaccurate predictions (Oussous *et al.*, 2018).

In summary, statistical and probability theory provide powerful tools for analyzing and managing big data in banking, but their application requires careful consideration of assumptions, data quality, and model uncertainties to ensure robust and reliable results.

3.3 Machine learning and data mining theory

Machine learning and data mining are vital components of big data management in the banking sector, enabling financial institutions to extract valuable insights from large volumes of data (Akter & Wamba, 2016). Machine learning and data mining propose that by applying algorithms and statistical techniques to big data, banks can uncover patterns, trends, and relationships that can inform decision-making processes, improve risk management, and enhance customer experience. These approaches involve training models on historical data to learn patterns and relationships, which can then be used to make predictions or identify anomalies in new data. Machine learning algorithms can adapt to changing data patterns and dynamics, allowing banks to continuously improve their models and strategies over time (Akter & Wamba, 2016).

Machine learning models can leverage large datasets to make accurate predictions about customer behavior, credit risk, market trends, and other key variables, helping banks make more informed decisions. Data mining and machine learning techniques are scalable to handle large volumes of data, making them well-suited for big data management in banking (Zhou *et al*., 2017).

The theory assumes that historical data is representative of future behavior and can be used to train predictive models and that the relationships and patterns discovered in the data are meaningful and generalizable to new data instances (Richins *et al.*, 2017). Machine learning models assumes that data is labeled or can be labeled for supervised learning tasks, although unsupervised and semi-supervised techniques can also be used in certain scenarios.

Machine learning models may capture noise or spurious patterns in the data, leading to overfitting and poor generalization performance on unseen data (Akter & Wamba, 2016). Machine learning algorithms are sensitive to the quality of input data, and preprocessing steps such as cleaning, normalization, and feature engineering are often required to ensure optimal performance. Some machine learning models, such as deep neural networks, are inherently complex and may lack interpretability, making it challenging for banks to understand and trust the decisions made by these models (Richins *et al.*, 2017).

Banks should prioritize data quality initiatives and establish robust data preprocessing pipelines to ensure that input data is clean, accurate, and well-suited for machine learning tasks (Sun *et al.*, 2017). Banks should consider using simpler, more interpretable machine learning models whenever possible, especially in regulatory or high-stakes decision-making contexts where model transparency and accountability are important. Banks should encourage a culture of experimentation and continuous improvement, where teams are empowered to explore new machine learning techniques, evaluate their effectiveness, and iterate on existing models and strategies (Zhao *et al.*, 2017).

Table 1: Listed of all commercial banks in Nigerian banking sector

S/No	Name of Bank	Classification according to Central Bank of Nigeria
	Commercial Bank	
1	Access Bank Limited	International
2	Fidelity Bank Plc	International
3	First City Monument Bank Limited	International
4	First Bank Nigeria Limited	International
5	Guaranty Trust Bank Limited	International
6	United Bank of Africa Plc	International
7	Zenith Bank Plc	International
	Commercial Bank	
1	Citibank Nigeria Limited	National
2	Ecobank Nigeria Limited	National
3	Heritage Bank Plc	National
4	Globus Bank Limited	National
5	Keystone Bank Limited	National

6	Polaris Bank Limited	National
7	Stanbic IBTC Bank Limited	National
8	Standard Chartered Bank Limited	National
9	Sterling Bank Limited	National
10	Titan Trust Bank Limited	National
11	Union Bank of Nigeria Plc	National
12	Unity Bank Plc	National
13	Wema Bank Plc	National
14	Premium Trust Bank Limited	National
15	Optimus Bank Limited	National
	Commercial Bank	
1	Parallex Bank Limited	Regional
2	Providus Bank Limited	Regional
3	SunTrust Bank Limited	Regional
4	Signature Bank Limited	Regional

Source: Central Bank of Nigeria (CBN), 2024

4.0 General outlooks and conditions of big data management in banking sector in Nigeria

Managing big data in a developing economy like Nigeria involves leveraging its strengths, addressing weaknesses, capitalizing on opportunities, and mitigating threats.

4.1 SWOT analysis

Strengths:

1. Increasing internet penetration: More people have access to the internet, generating more data.

2. Growing tech-savvy workforce: A rising number of skilled professionals can handle big data analytics.

3. Government support: Initiatives to promote technology and innovation can boost big data management.

Weaknesses:

1. Infrastructure challenges: Inadequate internet connectivity and power supply hinder data collection and processing.

2. Limited data literacy: Many lack the skills to analyze and interpret big data effectively.

3. Data privacy concerns: Weak regulations and enforcement expose sensitive data to risks.

Opportunities:

1. Untapped market potential: Big data analytics can drive insights for businesses across various sectors like agriculture, finance, and healthcare.

2. Foreign investment: International firms may invest in Nigeria's big data infrastructure, driving innovation and economic growth.

3. Public-private partnerships: Collaborations between government and private sectors can enhance data management capabilities.

Threats:

1. Security risks: Cyber threats and data breaches pose significant challenges to managing big data securely.

2. Economic instability: Fluctuations in the economy can impact investment in big data infrastructure and talent development.

3. Competition from other markets: Nigeria may face competition from other emerging economies in attracting investments and talent for big data management.

Analyzing large volumes of data allows banks to gain deep insights into customer behavior, preferences, and needs. This understanding enables banks to offer personalized services and products tailored to individual customers, leading to higher customer satisfaction and retention.

Big data analytics helps banks identify and mitigate various risks, including credit, market, operational, and regulatory risks (Richins *et al.*, 2017). By analyzing historical data and real-time transactions, banks can detect fraud, predict default rates, and assess the creditworthiness of borrowers more accurately.

Big data technologies enable banks to streamline their operations and improve efficiency. By automating routine tasks, optimizing processes, and identifying bottlenecks, banks can reduce costs, enhance productivity, and deliver services more quickly and effectively (Sookhak *et al.*, 2017).

Analyzing big data allows banks to identify emerging market trends and customer demands, enabling them to develop innovative products and services. This agility in product development can give banks a competitive edge in the market and attract new customers (Boone *et al.*, 2017).

Compliance with regulatory requirements is a significant challenge for banks, especially in developing economies where regulations may be evolving rapidly. Big data analytics helps banks ensure compliance by monitoring transactions, detecting suspicious activities, and generating reports required by regulatory authorities.

Implementing big data management in a developing economy like Nigeria comes with several limitations: Limited access to reliable internet connectivity, power outages, and inadequate technology infrastructure can hinder the effective collection, storage, and processing of large volumes of data. Inconsistent data quality due to factors like manual data entry errors, lack of standardized data formats, and incomplete data sets can undermine the accuracy and reliability of big data analytics. Shortage of skilled professionals with expertise in big data technologies and analytics can impede the effective utilization of available data for decision-making and innovation.

Unclear data protection and privacy regulations, coupled with concerns about data security and governance, can pose significant challenges to implementing big data initiatives in compliance with local laws and international standards (Seddon & Currie, 2017). Limited budget allocations for investing in advanced technology infrastructure, data management systems, and training programs can restrict the scalability and sustainability of big data projects in developing economies (Boone *et al.*, 2017). Resistance to change, hierarchical organizational structures, and a lack of data-driven culture within businesses and government agencies can slow down the adoption and integration of big data solutions.

Addressing these limitations requires a multi-faceted approach involving investment in infrastructure development, capacity building, regulatory reforms, and fostering a culture of data-driven decision-making (Côrte-Real *et al.*, 2019).

5.0 **Conclusion and recommendation**

In order to leverage on robust big data management technologies, Nigerian banking sector should establish a robust data governance framework to ensure data quality, security, and compliance with regulations such as the Nigeria Data Protection Regulation (NDPR). This framework should define roles, responsibilities, policies, and procedures for managing data throughout its lifecycle. Banks should upgrade data infrastructure to handle the volume, variety, and velocity of big data. This may involve investing in scalable storage solutions, cloud computing resources, and data integration platforms.

Banking sector players should leverage advanced analytics techniques such as predictive modeling, machine learning, and artificial intelligence to extract insights from big data. These insights can be used for customer segmentation, risk management, fraud detection, and personalized marketing. There is also need to strengthen cybersecurity measures to protect sensitive customer data from cyber threats. This includes implementing encryption, access controls, and monitoring tools to detect and respond to security incidents. Furthermore, bankers should ensure compliance with data privacy regulations such as the General Data Protection Regulation (GDPR) and the NDPR. This involves obtaining consent for data collection and processing, providing transparency about data usage, and honoring data subject rights.

In conclusion, enhancing effective and efficient big data management in the banking sector of a developing economy like Nigeria involves several key recommendations: Commercial bank managements should:

- Build robust data infrastructure capable of handling large volumes of data securely and efficiently.
- Implement strong data governance policies and procedures to ensure data quality, security, and compliance with regulations.
- Prioritize customer privacy and data protection to maintain trust and confidence in the banking sector.
- Implement scalable solutions that can accommodate the growing volume and complexity of data in the banking sector.
- Stay abreast of regulatory requirements and ensure compliance with data protection and privacy regulations.
- Integrate data from various sources within the bank and external sources to gain a comprehensive view of customers, operations, and market trends.
- Continuously evaluate and improve big data management processes and technologies to stay ahead of evolving challenges and opportunities.

By implementing these recommendations, commercial banks in Nigeria can harness the power of big data to drive innovation, improve decision-making, and enhance competitiveness in the digital era.

AUTHORS' CONTRIBUTIONS

This work was carried out by the authors. The three authors read and approved the final manuscript.

COMPETING INTERESTS

Authors declared that no competing interests exist.

FUNDING

No financial support or external funder for this study.

References

Ahamad, D., Akhtar, M. & Hameed, S. A. (2019). A review and analysis of big data and mapreduce. *International Journal of Advance Trends Computer Science Engineering,8*(1),1-3.

Ahmad, J., Muhammad, K., Lloret, J., & Baik, S. W. (2018). Efficient conversion of deep features to compact binary codes using fourier decomposition for multimedia big data. *IEEE Transactions on Industrial Informatics,14*(7),3205-3215.

Akter, S., & Wamba, S. F. (2016). Big data analytics in E-commerce: A systematic review and agenda for future research. *Electronic Markets,26*(2),173-194.

Alzakholi, O., Shukur, H., Zebari, R., Abas, S., & Sadeeq, M. (2020). Comparison among cloud technologies and cloud performance. *Journal of Applied Sciences, Technology & Trends,1*(2),40-47.

Boone, C. A., Skipper, J. B., & Hazen, B. T. (2017). A framework for investigating the role of big data in service parts management. *Journal of Cleaner Production,153*, 687-691.

Côrte-Real, N., Ruivo, P., Oliveira, T., & Popovič, A. (2019). Unlocking the drivers of big data analytics value in firms. *Journal of Business Research,97*,160-173.

Mhlanga, D. (2020). Industry 4.0 in finance: The impact of artificial intelligence (AI) on digital financial inclusion. *International Journal of Financial Studies*,*8*,45-55.

Mhlanga, D. (2021). Financial inclusion in emerging economies: The application of machine learning and artificial intelligence in credit risk assessment. *International Journal of Financial Studies*,*9*,39-47.

Mhlanga, D. (2022). The role of artificial intelligence and machine learning amid the COVID-19 pandemic: What lessons are we learning on 4IR and the sustainable development goals. *International Journal of Environmental Research and Public Health*,*19*,1879-1888.

Oussous, A., Benjelloun, F. Z., Lahcen, A. A., & Belfkih,S. (2018). Big data technologies: A survey. *Journal of King Saud University-Computer and Information Sciences,30*(4),431-448.

Richins, G., Stapleton, A., Stratopoulos, T. C., & Wong, C. (2017). Big data analytics: Opportunity or threat for the accounting profession?. *Journal of Information Systems,31*(3), 63-79.

Seddon, J. J., & Currie, W. L. (2017). A model for unpacking big data analytics in high-frequency trading. *Journal of Business Research,70*,300-307.

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research,70*,263-286.

Sookhak, M., Gani, A., Khan, M. K., & Buyya, R. (2017). Dynamic remote data auditing for securing big data storage in cloud computing. *Information Sciences,380*,101-116.

Sun, F., Huang, G. B., Wu, Q. J., Song, S., & Wunsch, D. C. II. (2017). Efficient and rapid machine learning algorithms for big data and dynamic varying systems. *IEEE Transactions on Systems, Man, and Cybernetics Systems,47*(10), 2625-2626.

Zhao, R., Liu, Y., Zhang, N., & Huang, T. (2017). An optimization model for green supply chain management by using a big data analytic approach. *Journal of Cleaner Production,142*,1085-1097.

Zhou, L., Pan, S., Wang, J., & Vasilakos, A. V. (2017). Machine learning on big data: Opportunities and challenges. *Neurocomputing,237*,350-361.