University for the Creative Arts

BERLIN SCHOOL OF BUSINESS & INNOVATION

Essay Title: An essay on Big Data Analytics in a banking system

Name: Sachin Mehta

Date: 2023

Statement of compliance with academic ethics and the avoidance of plagiarism

I honestly declare that this essay is entirely my work and none of its part has been copied from printed or electronic sources, translated from foreign sources, or reproduced from essays of other researchers or students. Wherever I have been based on ideas or other people's texts I declare it through the good use of references following academic ethics.

(In the case that is proved that part of the essay does not constitute an original work, but a copy of an already published essay or from another source, the student will be expelled permanently from the postgraduate program).

Name and Surname (Capital letters):

SACHIN MEHTA

Date: 2023/10/06

Executive summary

Big Data analytics has changed the game in the banking sector, which is continually developing. An in-depth analysis of the underlying traits of big data analytics, the benefits of file systems, and how they enable learning and decision-making are covered in this paper. With an emphasis on data retrieval simplicity, data format flexibility, and real-time processing capabilities, it specifically studies how banks extract insights from enormous data repositories. Successful applications, such as fraud detection, credit risk analysis, and client segmentation, are also highlighted in the paper. Big Data analytics is altering banking operations, improving effectiveness, controlling risks, and upgrading client experiences, eventually influencing the direction of the financial industry.

TABLE OF CONTENTS

INTRODUCTION

Concept of Big Data and Its Significance

Big Data is a term used to describe the enormous amount, velocity, diversity, validity, and value of data produced within the sector. Large volumes of structured and unstructured data, including client transactions, account information, market movements, social media sentiment, and more, are gathered every day by the banking industry (Nguyen et al., 2019:29). Big Data analytics' importance to the banking industry cannot be emphasized. It equips financial institutions with the tools they need to reduce risks, improve client experiences, and make data-driven choices. Big Data analytics, for instance, may identify possible investment possibilities by crunching market data, assist in fraud detection by examining transaction patterns, and customize financial services depending on user behavior (Izraelevitz et al., 2019:58).

Overview of the banking industry's increasing reliance on Big Data Analytics

The banking sector is going finished a significant transition, which is being characterized by an increasing requirement on big data analytics (Izraelevitz et al., 2019:58). This change is a result of the understanding that data is a priceless reserve that can be used to improvement an advantage over contestants. Banks are increasingly using huge capacities of client information, transaction history, and market data to help guide decision-making. Banks may deliver more specialized and beleaguered services by using big data analytics to better understand client preferences and behavior. It is crucial for risk evaluation, fraud detection, and devotion to strict regulatory standards (Nguyen et al., 2019:29).

Task 1: Big Data in the Banking System

1.1 Explanation of Big Data Analytics in Banking

Big data analytics refers to the systematic procedure of congregation, processing, analyzing, and interpretation huge and varied information produced within the financial sector. This strategy makes use of cutting-edge tools and logical methods to glean significant trends, patterns, and insights from the huge quantity of data that banks have amassed (Dai et al., 2019:55).

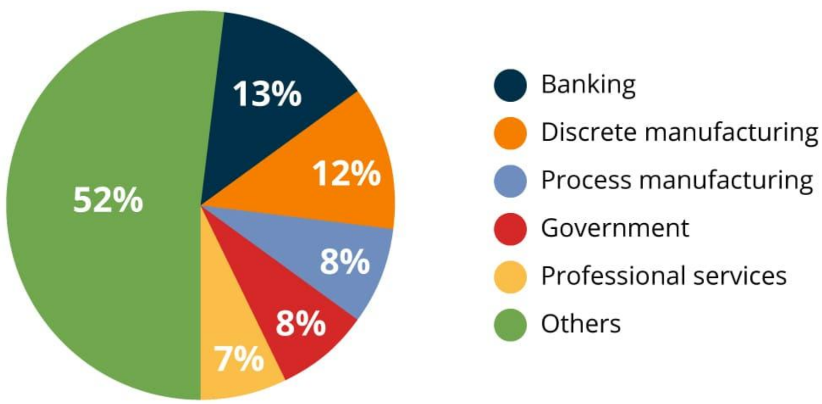


Figure 1: Explanation of Big Data Analytics in Banking

Source: (Itsvit., 2023, Big Data in the Banking System:1)

Customer Insights and Risk Management

To get a thorough knowledge of consumer behaviors, welfares, and demands, banks utilize big data analytics. They can personalize offers, improve consumer experiences, and create focused marketing tactics by investigative transaction histories, online interactions, and social media sentiment. Banks may examine and reduce a variety of risks, including credit, market, and operational risks, with the use of big data analytics. It involves keeping track of shifts in the state of the economy, spotting indiscretions in transaction data, and foreseeing possible problems before they become serious (Nguyen et al., 2019:29).

Fraud Detection and Compliance and Regulatory Reporting

Banks use Big Data analytics to find unforeseen trends in transaction data that indicate fraudulent activity. This proactive method shields clients against unauthorized transactions and helps avoid financial losses. Banks must follow stringent reporting and compliance requirements, according to regulatory agencies. By automating these procedures, big data analytics lowers the risk of non-compliance while promising correct and timely filings (Dai et al., 2019:55).

Operational Efficiency

Big Data analytics are strategically used by banks to improve the efficacy and efficiency of their internal operations. Banks may distribute cash and resources where they are most required by optimizing resource allocation by analyzing consumer foot traffic and transaction data (Nizamuddin et al., 2019:183-197). Identifying underperforming or overstaffed branches and changing their locations or services as necessary constitutes branch network optimization (Saiz-Rubio and Rovira-Más, 2020:59). Data-driven insights are useful for workforce management because they help banks better schedule employees, match skill sets to jobs, and cut down on wasteful labor expenditures. Banks may achieve considerable cost reductions and operational enhancements while maintaining high levels of service quality by using Big Data analytics in these areas (Saiz-Rubio and Rovira-Más, 2020:59).

1.2 Specific examples or case study of Big Data in Banking

The crucial function of fraud detection played by big data in banking is an effective illustration. Banks routinely track millions of transactions every day, and the amount of data they collect makes it virtually hard to manually identify fraudulent activity. Big Data analytics stand out as a game-changer in this situation. One case study that serves as an example concerns a well-known bank that used cutting-edge analytics to fight fraud. To examine consumer behavior and transaction trends, the bank used machine learning algorithms. The technology detected irregularities and reported possible fraudulent activity in real-time by looking at prior data. For instance, the system raised an alarm when it saw an unexpected increase in ATM withdrawals from several locations using the same account. In addition, it detected a rapid shift in purchasing patterns that included sizable purchases made abroad. The technology compared this data to the customer's prior trip experiences and usual buying habits (Nizamuddin et al., 2019:183-197).

1.3 Fundamental characteristics of Big Data in this scenario

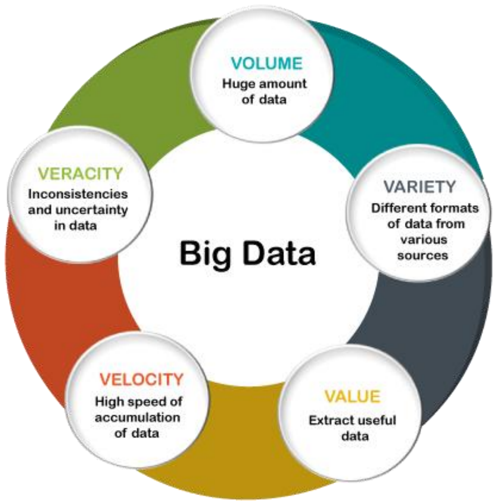


Figure 2: Fundamental characteristics of Big Data in this scenario

Source: (Javatpoint, 2023, Big Data Characteristics:2)

Volume

The amount of data used in banking is enormous. Innumerable activities, including deposits, withdrawals, internet transfers, credit card use, and customer interactions, create a massive quantity of data for banks every day. This information includes historical records that span years' worth of financial activities. The amount of data is astonishing, necessitating specialized systems and instruments for storage (Yaqoob et al., 2021:1-16).

Velocity

In the financial industry, data moves at an astounding rate. Real-time transactions and round-the-clock client service are provided via a variety of channels, including Internet banking, mobile applications, ATMs, and in-person branch visits. Because of the speed at which data is generated, it must be processed and analyzed right away to make timely choices, particularly in fields like fraud detection where a delay might lead to large losses (Nizamuddin et al., 2019:183-197).

Variety

Various formats are available for banking data. Account balances, transaction histories, and client information are all examples of structured data that are orderly and organized. Emails, social media interactions, client comments, and even contact center recordings are all examples of unstructured data, in contrast. To fully understand consumer behavior and market trends, banks must use both structured and unstructured data (Yaqoob et al., 2021:1-16).

Veracity

In the banking industry, maintaining data accuracy and quality is a key concern. Data input mistakes may happen, and combining data from several sources might lead to discrepancies. Verifying data accuracy is crucial since faulty data may result in poor judgments and operational inefficiencies. To overcome these issues, data validation and purification are essential tasks (Badidi, Mahrez and Sabir, 2020:190).

Value

Big Data analytics provides banks with significant value. By optimizing procedures like risk assessment and resource allocation, it improves operational efficiency. It makes it possible to provide better customer service by tailoring products and anticipating demands. Additionally, it helps to increase income via enhanced cross-selling possibilities and the discovery of lucrative market niches. Big Data analytics equips banks to make data-driven choices that boost revenue, lower risks, and strengthen their position as market leaders in the financial sector (Urbinati et al., 2019:21-36).

Task 2: Database Management Systems vs. File Systems

2.1 Database Management Systems (DBMS) and their role in data storage and management

Data storage, organization, retrieval, and administration inside organizations are all fundamentally facilitated by database management systems (DBMS), which are software programs. Their main objective is to provide a systematic and effective method of storing, accessing, and manipulating massive amounts of data. Here is a detailed explanation of DBMS and its function (Badidi, Mahrez and Sabir, 2020:190).

Data Storage and Data Retrieval

A centralized data store is provided by DBMS. It establishes linkages between various kinds of information while organizing data into structured tables. Data storage and retrieval are made easier by this organized format, which also ensures data consistency and integrity. Through standardized query languages like SQL (Structured Query Language), DBMS allows users to quickly obtain data. Data accessibility is improved by the ability of users to select criteria to rapidly and reliably obtain certain data subsets (Urbinati et al., 2019:21-36).

Data Security and Data Integrity

DBMS systems include strong security controls, such as user authentication and authorization, to manage data access. This protects sensitive information by ensuring that only authorized workers may access, alter, or delete data. Constraints on data integrity, such as unique keys, referential integrity, and data validation criteria, are enforced by DBMS. These limitations aid in preserving data consistency and accuracy by avoiding mistakes and data corruption (Yang et al., 2020:45).

Concurrency Control and Scalability and Performance

In contexts with many users, DBMS controls concurrent access to data, avoiding conflicts and guaranteeing data consistency. For real-time systems to ensure data integrity, this characteristic is essential. Large datasets may be effectively handled by DBMS systems. They provide tools for indexing, caching, and query optimization, guaranteeing quick data retrieval even as data quantities increase (Urbinati et al., 2019:21-36).

Backup and Recovery and Data Analysis

In the event of system failures or data loss incidents, DBMS contains functionality for data backup and recovery, enabling organizations to restore data. Several DBMS systems provide reporting and data analysis features. These technologies support decision-making by enabling users to get insights from the data that has been saved (Yang et al., 2020:45).

2.2 File Systems and how they differ from DBMS

File Systems

Structure and Access

Data is kept in hierarchically organized files and directories by file systems. Data often lacks a set format and is unstructured or semi-structured. File systems provide fundamental data retrieval and access. There is no standardized query language for sophisticated data retrieval, and users directly interact with files (Chang et al., 2020:120).

Data Redundancy Scalability, and Concurrency

Since the same data may reside in several files, data redundancy is a typical problem in file systems. Inconsistency and greater storage needs may result from this. File system scaling may be difficult. The management and upkeep of file and directory organization becomes more difficult as data volume rises. Data integrity problems may arise in multi-user situations because file systems often have limitations regarding the simultaneous access to data by numerous users (Yang et al., 2020:45).

Database Management Systems (DBMS)

Structure

In organized tables with preset schemas, DBMS stores data. Rows and columns are used to organize data, guaranteeing its uniformity and integrity (Chang et al., 2020:120).

Access and Data Redundancy

The ability to access and retrieve complex data using SQL or other query languages is provided by DBMS. Complex searches may be run by users to get certain data subsets. By using normalization methods, DBMS reduces data redundancy, which lowers storage needs and ensures data consistency (Sarker et al., 2020:7).

Scalability and Concurrency

Large datasets may be effectively handled by DBMS systems. To sustain performance as data accumulates, they offer indexing, optimization, and clustering. Concurrent access control techniques are included in DBMS systems to protect data integrity in multi-user settings. Large datasets may be effectively handled by DBMS systems. To sustain performance as data accumulates, they offer indexing, optimization, and clustering. Concurrent access control techniques are included in DBMS systems to protect data integrity in multi-user settings (Chang et al., 2020:120).

2.3 File Systems appropriateness for most Big Data analytic approaches in the banking industry

Due to their adaptability and scalability, file systems are better suited for many Big Data analytic methodologies in the banking sector. Data in banking may be found in a variety of forms, including unstructured data like emails and social media interactions and structured data like transaction records (Wang et al., 2020:393-405). Daily data production in the banking industry is enormous, and File Systems can readily grow to handle this massive data flow without the need for laborious database settings or schema changes. The processing of the expanding number of transactional and customer data depends on this scalability (Sarker et al., 2020:7).

2.4 Scalability, flexibility, and cost-effectiveness as key reasons for choosing File Systems

1. Scalability

Due to their adaptability and scalability, file systems are better suited for many Big Data analytical methodologies used in the banking sector. Data in the banking industry might be organized, like transaction records, or unstructured, like emails and social media interactions. Without the limitations of inflexible database schemas, file systems may manage this variety of data with efficiency, making them suited for managing the many data types used in the financial industry (Sarker et al., 2020:7). File Systems may quickly expand to accommodate this massive data flow without the need for complicated database settings or schema changes, which is necessary given the enormous amount of data that the banking industry produces every day. For processing the expanding number of transactional and consumer data, this scalability is essential (Wang et al., 2020:393-405).

2. Flexibility

The versatility of file systems in footings of data kinds and formats is quite substantial. Data is used in banking in a diversity of ways, such as structured transaction annals, unstructured customer reviews, and semi-structured market info. File Systems can achieve this variety of data without the strict plan restrictions common to DBMS. This flexibility is chiefly useful in Big Data analytics, where a typical trouble is the diversity of the data (Helbing et al., 2018:49).

3. Cost-Effectiveness

It may be expensive to implement and uphold a DBMS, both in terms of certifying costs and hardware supplies. File Systems, on the other hand, are often more inexpensive. They need less money upfront and are often humbler to operate. In the banking sector, where cost discount is a top focus, file systems deliver a desirable solution for managing large data amounts while limiting operating costs (Wang et al., 2020:393-405).

Task 3: Learning from Big Data with File Systems

3.1 Extraction of insights and learning from Big Data using File Systems

Banks may use file systems to use Big Data to their advantage by following a few essential procedures. They may use data intake methods to gather and store enormous volumes of structured and unstructured data from a variety of sources, including transaction records, customer interactions, and market data (Helbing et al., 2018:49). These files may be easily processed by powerful data analytics tools and frameworks, allowing banks to conduct in-depth investigations, spot trends, and find hidden relationships in the data. Predictive models for fraud detection, risk assessment, and consumer behavior analysis may be created using machine learning techniques (Reddy et al., 2020:55).

3.2 Attributes and properties of File Systems that support the learning and decision-making processes

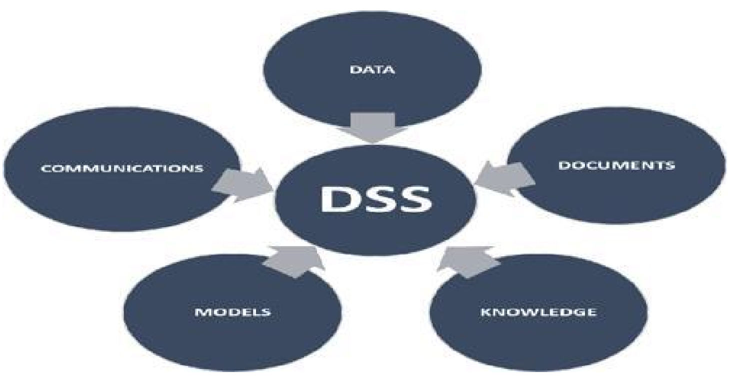


Figure 3: File Systems that support the learning and decision-making processes

Source: (Qsstudy, 2023, Characteristics of Decision Support System (DSS):49)

1. Ease of Data Retrieval and Storage

File systems provide a simple method for storing and retrieving data. Intuitive file organization makes it simple to find and retrieve particular data when required (Saadoon et al., 2022:101). This characteristic guarantees that analysts may quickly and effectively collect pertinent data for analysis in the banking industry, where rapid access to previous transaction records or client information is essential. Additionally, File Systems are extremely scalable as data grows over time, enabling institutions to store enormous volumes of data without requiring laborious setup modifications. Because of its scalability, historical data is always available for trend analysis and decision-making (Saadoon et al., 2022:101).

2. Ability to Handle Various Data Formats

Banking data often comes in a variety of forms, from unstructured data like emails and social media interactions to structured data like client profiles. File systems don't need a set structure, therefore they can handle this range of data with ease (Helbing et al., 2018:49). Banks can store, handle, and analyze a variety of data types with this flexibility, which makes it simpler to get insightful information. File Systems can effectively manage any data format, whether it be structured transaction records or unstructured consumer input, providing thorough analytics (Reddy et al., 2020:55).

3. Real-Time Processing Capabilities

Real-time data processing is essential in the banking sector for spotting fraud, controlling risk, and quickly reacting to market developments. Banks can consume and analyze data streams as they come in thanks to file systems' real-time processing capabilities (Reddy et al., 2020:55). As a result, banks are better equipped to make decisions on time and respond quickly to urgent situations or opportunities, increasing their overall operational effectiveness and market competitiveness. Banks may use their Big Data assets to the fullest extent possible for data-driven, informed choices because of their capacity to handle data in real-time (Saadoon et al., 2022:101).

3.3 Role of data mining, machine learning, and predictive analytics in leveraging Big Data for decision-making in banking

1. Data Mining

Banks may find significant patterns and insights from enormous datasets by using data mining methods. Banks may find hidden connections and anomalies by examining historical transaction data, client interactions, and market movements. Data mining, for instance, might show anomalous spending patterns that point to fraud or client preferences for tailored services. These insights enable banks to make wise choices, improve client experiences, and efficiently manage risks (Mikalef et al., 2019:19).

2. Machine Learning

Banks may create prediction models using machine learning algorithms for a variety of uses, including credit risk assessment and fraud detection. These algorithms can examine enormous volumes of data, spot trends, and make predictions based on past trends. For instance, real-time detection of questionable activity is possible using machine learning models, as can the evaluation of a borrower's creditworthiness. Banks can decrease human error, improve accuracy, and speed up decision-making by automating these operations (N. et al., 2022:55).

3. Predictive Analytics

To predict future occurrences or trends, predictive analytics makes use of both historical and current data. This method helps forecast client behavior, market volatility, and possible financial concerns in the banking industry. By employing predictive analytics, banks may build methods to minimize prospective problems, optimize resource allocation, and proactive adapt their services, eventually resulting in better financial planning and better decision-making (Mikalef et al., 2019:19).

3.4 Examples of successful applications of Big Data analytics in banking

Fraud Detection

Banks can quickly spot strange transaction patterns and abnormalities using big data analytics, identifying probable fraud cases in real-time. Machine learning algorithms may identify fraudulent activity like unauthorized credit card use or account takeovers by analyzing enormous quantities of transaction data. Banks can safeguard their consumers from fraud thanks to this proactive strategy, which also helps them avoid financial losses (N. et al., 2022:55).

Credit Risk Assessment

To evaluate a borrower's creditworthiness, big data analytics uses a variety of data sources, such as financial history, social behavior, and market movements. Banks can improve loan approval procedures, lower the likelihood of defaults, and make more precise lending choices by analyzing these characteristics (Mikalef et al., 2019:19).

Customer Segmentation

Banks segment their client base based on behavior and preferences using data-driven insights. This makes it possible to customize marketing initiatives and services. For instance, a bank may recognize a group of wealthy people and provide investment possibilities only for them. By providing individualized experiences, this strategy improves client happiness and retention (N. et al., 2022:55).

CONCLUSIONS

Big Data analytics has become a game-changing force in the banking sector, completely changing how financial organizations function and make decisions. Several important elements have supported this technological change. Big Data analytics' growing use in the banking industry is fueled by its capacity to draw important conclusions from enormous and varied datasets. To fully realize the potential of the volume, velocity, and diversity of data created by banking transactions and client interactions, new analytics tools and methodologies are needed. File systems and database management systems (DBMS) provide separate but complementary functions in managing and storing data. File Systems provide flexibility and scalability in contrast to DBMS's capacity to store structured data, making them more suitable for many Big Data analysis techniques used in the banking industry.

BIBLIOGRAPHY

Badidi, E., Mahrez, Z. and Sabir, E. (2020). Fog Computing for Smart Cities' Big Data Management and Analytics: A Review. Future Internet, 12(11), p.190. doi:https://doi.org/10.3390/fi12110190.

Chang, V., Baudier, P., Zhang, H., Xu, Q., Zhang, J. and Arami, M. (2020). How Blockchain can impact financial services – The overview, challenges and recommendations from expert interviewees. Technological Forecasting and Social Change, [online] 158(1), p.120166. doi:https://doi.org/10.1016/j.techfore.2020.120166.

Coronel, C. and Morris, S.A. (2022). Database Systems: Design, Implementation and Management. [online] Google Books. Cengage Learning. Available at: https://books.google.com/books?hl=en&lr=&id=AKOIEAAAQBAJ&oi=fnd&pg=PP1&dq=Database+Management+Systems+vs.+File+Systems&ots=QAaH3pSxxX&sig=vRqf8fgmTozTWjlgovtAAbrfOjo [Accessed 22 Sep. 2023].

Dai, H.-N., Wang, H., Xu, G., Wan, J. and Imran, M. (2019). Big data analytics for manufacturing internet of things: opportunities, challenges and enabling technologies. Enterprise Information Systems, 14(9-10), pp.1–25. doi:https://doi.org/10.1080/17517575.2019.1633689.

Helbing, D., Frey, B.S., Gigerenzer, G., Hafen, E., Hagner, M., Hofstetter, Y., van den Hoven, J., Zicari, R.V. and Zwitter, A. (2018). Will Democracy Survive Big Data and Artificial Intelligence? Towards Digital Enlightenment, [online] pp.73–98. doi:https://doi.org/10.1007/978-3-319-90869-4_7.

Izraelevitz, J., Yang, J., Zhang, L., Kim, J., Liu, X., Memaripour, A., Soh, Y.J., Wang, Z., Xu, Y., Dulloor, S.R., Zhao, J. and Swanson, S. (2019). Basic Performance Measurements of the Intel Optane DC Persistent Memory Module. arXiv:1903.05714 [cs]. [online] Available at: https://arxiv.org/abs/1903.05714.

Mikalef, P., Krogstie, J., Pappas, I.O. and Pavlou, P. (2019). Exploring the relationship between big data analytics capability and competitive performance: The mediating roles of dynamic and operational capabilities. Information & Management, [online] 57(2). doi:https://doi.org/10.1016/j.im.2019.05.004.

N., D., Pham, Q.-V., Nguyen, D.C., Bhattacharya, S., Prabadevi, B., Gadekallu, T.R., Maddikunta, P.K.R., Fang, F. and Pathirana, P.N. (2022). A survey on blockchain for big data: Approaches, opportunities, and future directions. Future Generation Computer Systems, 131. doi:https://doi.org/10.1016/j.future.2022.01.017.

Nguyen, D.C., Pathirana, P.N., Ding, M. and Seneviratne, A. (2019). Blockchain for Secure EHRs Sharing of Mobile Cloud Based E-Health Systems. IEEE Access, 7, pp.66792–66806. doi:https://doi.org/10.1109/access.2019.2917555.

Nizamuddin, N., Salah, K., Ajmal Azad, M., Arshad, J. and Rehman, M.H. (2019). Decentralized document version control using ethereum blockchain and IPFS. Computers & Electrical Engineering, [online] 76, pp.183–197. doi:https://doi.org/10.1016/j.compeleceng.2019.03.014.

Reddy, G.T., Reddy, M.P.K., Lakshmanna, K., Kaluri, R., Rajput, D.S., Srivastava, G. and Baker, T. (2020). Analysis of Dimensionality Reduction Techniques on Big Data. IEEE Access, [online] 8, pp.54776–54788. doi:https://doi.org/10.1109/ACCESS.2020.2980942.

Saadoon, M., Ab. Hamid, S.H., Sofian, H., Altarturi, H.H.M., Azizul, Z.H. and Nasuha, N. (2022). Fault tolerance in big data storage and processing systems: A review on challenges and solutions. Ain Shams Engineering Journal, [online] 13(2), p.101538. doi:https://doi.org/10.1016/j.asej.2021.06.024.

Saiz-Rubio, V. and Rovira-Más, F. (2020). From Smart Farming towards Agriculture 5.0: A Review on Crop Data Management. Agronomy, 10(2), p.207. doi:https://doi.org/10.3390/agronomy10020207.

Sarker, I.H., Kayes, A.S.M., Badsha, S., Alqahtani, H., Watters, P. and Ng, A. (2020). Cybersecurity data science: an overview from machine learning perspective. Journal of Big Data, 7(1). doi:https://doi.org/10.1186/s40537-020-00318-5.

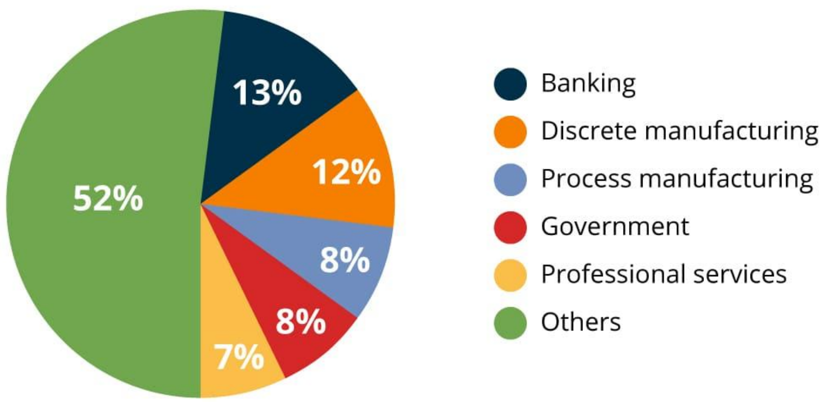
Urbinati, A., Bogers, M., Chiesa, V. and Frattini, F. (2019). Creating and capturing value from Big Data: A multiple-case study analysis of provider companies. Technovation, [online] 84-85, pp.21–36. doi:https://doi.org/10.1016/j.technovation.2018.07.004.

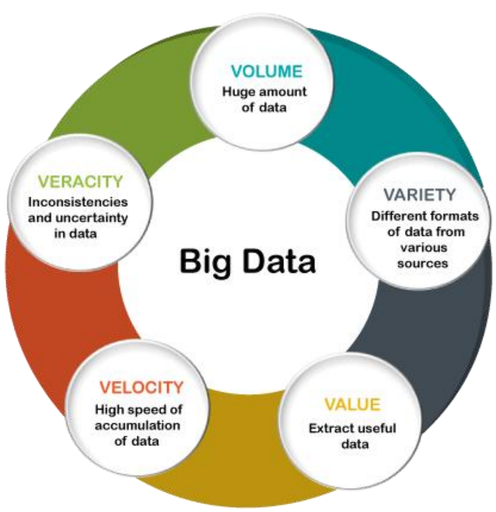
Wang, J., Yang, Y., Wang, T., Sherratt, R.S. and Zhang, J. (2020). Big Data Service Architecture: A Survey. Journal of Internet Technology, [online] 21(2), pp.393–405. Available at: https://jit.ndhu.edu.tw/article/view/2261.

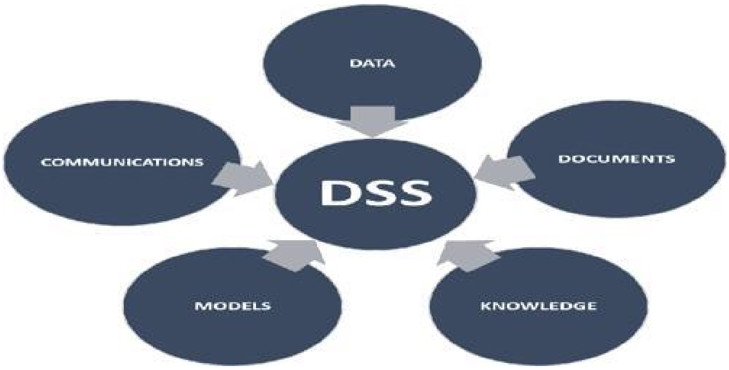
Yang, C., Lan, S., Wang, L., Shen, W. and Huang, G.G.Q. (2020). Big Data Driven Edge-Cloud Collaboration Architecture for Cloud Manufacturing: A Software Defined Perspective. IEEE Access, 8, pp.45938–45950. doi:https://doi.org/10.1109/access.2020.2977846.

Yaqoob, I., Salah, K., Jayaraman, R. and Al-Hammadi, Y. (2021). Blockchain for Healthcare Data management: opportunities, challenges, and Future Recommendations. Neural Computing and Applications, 34, pp.1–16. doi:https://doi.org/10.1007/s00521-020-05519-w.

APPENDIX


13%
12%
8%
8%
7%
52%
Banking
Discrete manufacturing
Process manufacturing
Government
Professional services
Others


VOLUME
Huge amount of data
VARIETY
Different formats of data from various sources
VALUE
Extract useful data
VELOCITY
High speed of accumulation of data
VERACITY
Inconsistencies and uncertainty in data
Big Data


DATA
COMMUNICATIONS
DSS
DOCUMENTS
MODELS
KNOWLEDGE