

Advancements of Big Data in Banking

Sanjay Rawal

Big data refers to the significant amount of structured and unstructured data that organisations gather and analyse to obtain knowledge and improve decision-making. In the banking sector, big data provides information on client transactions, market trends and economic indicators. Banking institutions use big-data analytics to find patterns and trends, spot fraud and produce precise risk assessments. This helps banks improve their customer service, find new business prospects and make better financing and investment decisions. [1]

Using both personal and transactional information, banks can establish a 360-degree view of their customers to do the following:



Importance of big data in banking

The technology generates valuable insights for banks to make informed decisions, improve customer experience and streamline operations. It processes large amounts of data produced by exchanges with digital transaction, giving banks the ability to detect fraud, manage risk and gain a competitive advantage. Additionally, it helps banks personalise their offerings to meet unique customer needs. This results in increased efficiency and reduced cost, key factors for gaining a competitive advantage.

Implementing a big-data analytics strategy is in the best interest of financial institutions

The following are some of the key challenges:

- **Data complexity:** Banks generate large amounts of data from multiple sources (transactional, customer and market-related). This data can be complex and difficult to integrate, making it challenging to extract meaningful insights.
- **Data privacy and security:** With the increasing amount of sensitive customer data collected and analysed, banks should prioritise data privacy and security. This requires robust security measures such as encryption and access controls to protect customer data from unauthorised access and breaches.
- **Legacy systems:** Compared to other sectors, the banking sector innovates slowly. Most top banks run their businesses on IBM mainframes, and traditional banking institutions fall behind customer-focused firms in terms of agility. Legacy systems are unable to handle the burden, and the use of outmoded infrastructure to collect and store this data poses a risk to the system's overall stability.
- **Data governance:** Banks should establish clear data governance policies to ensure data is collected, processed and analysed in a consistent and compliant manner. This requires collaboration between the IT, legal and compliance departments to ensure data is used in a responsible and ethical manner.
- **Skilled workforce:** Implementation of big-data solutions requires a skilled workforce with a sound understanding of data analytics, machine learning and statistical modelling. This can be challenging for banks that do not have resources to hire or train data scientists and analysts.

Key benefits:

- **Customer insights:** Big data helps banks obtain detailed information on customer behaviour and preferences. This information can be used to develop targeted marketing strategies and improve customer engagement.
- **Fraud detection:** Big data helps detect and prevent fraud by identifying patterns and anomalies while analysing large volumes of transaction data.
- **Risk management:** By analysing large data volumes, banks can better understand and manage risk (credit, operational and market-related).
- **Improved operational efficiency:** Big data can help banks automate processes, reduce manual work and optimise their operations, resulting in improved efficiency and higher cost savings.
- **Better decision-making:** The technology provides a wealth of information that helps banks make informed decisions about products and services, risk management and investments.
- **Improved customer experience:** Big data helps banks discern customer needs and preferences, leading to improved customer experience and satisfaction.
- **Improved compliance:** Banks will be well prepared to meet regulatory requirements and improve compliance with anti-money-laundering and know-your-customer regulations.

A requisite in the banking sector:

- **Increasing data volumes:** The banking sector generates large amounts of data (e.g., transactional, customer interaction and market data) from various sources. These
-

volumes are often too high for traditional data processing methods to handle.

- **Need for real-time analysis:** Given the intense competition in the banking sector, banks should respond to changes and events in real time. Big data enables them to process large amounts of data in real time and make quick decisions.
- **Customer expectations:** Customers expect personalisation and convenience from banks. Big data provides the information needed to understand customer preferences and offer personalised services.
- **Compliance:** The banking sector is heavily regulated and must comply with a number of regulations. Big data can help banks meet requirements by providing the necessary data for compliance reporting and auditing.
- **Risk management:** Banks must manage risk to ensure the stability and profitability of the organisation. Big data provides the necessary information to identify and manage risks in real time.
- **Competitive advantage:** The ability to drive decision-making, improve customer experience and increase operational efficiency gives banks a competitive edge.

Toolkits for big-data processing in banks

Hadoop is a collection of open-source software utilities that use a network of computers to solve problems involving large amounts of data. It provides a software framework for distributed storage and processing of big data using the MapReduce programming model.

- **Data storage:** Hadoop provides a scalable and cost-effective solution for storing large amounts of data, both structured and unstructured.
 - **Data processing:** The MapReduce framework can process high data volumes in a parallel and distributed manner, reducing processing time and increasing efficiency.
-

- **Fraud detection:** Hadoop can store and analyse large amounts of transaction data to detect fraudulent activity.
- **Risk management:** It can analyse large amounts of [financial data](#) to identify potential risks and support decision-making.
- **Data security:** It provides a secure environment for storing and processing sensitive data, helping banks comply with regulations and protect customer information.

Spark is an in-memory data processing and analytics engine that can operate independently or in clusters controlled by Hadoop YARN, Mesos and Kubernetes. In addition to machine learning and graph processing, it can be utilised for batch and streaming applications. The following set of pre-installed modules and libraries offer a plethora of possibilities:

- **Data processing:** is used to process large amounts of data in a parallel and distributed manner, reducing the processing time and increasing efficiency.
- **Fraud detection:** It can process large volumes of transaction data to detect fraudulent activity.
- **Customer analytics:** It can process and analyse customer data, helping banks better understand their customer base and provide more personalised services.
- **Real-time streaming:** Spark can process real-time data streams, allowing banks to respond to events and changes in near real time via Spark Streaming.
- **Machine learning:** Spark's machine-learning libraries enable banks to build and deploy predictive models for fraud detection, customer analytics and risk management.

Airflow is an open-source platform for developing, scheduling and monitoring batch-oriented workflows. Its extensible Python framework helps build workflows connecting with all technologies, and a web interface manages the state of these workflows. Airflow is deployable in several ways, varying from a single process on a laptop to a distributed setup to support the biggest workflows. Its capabilities cover the following:

- **Data processing pipelines:** Automating and managing data processing tasks such as data extraction, transformation and loading
- **Workflow management:** Managing and scheduling complex workflows, including data processing and machine learning tasks
- **Job orchestration:** Coordinating and executing jobs, such as data backups and updates, across multiple systems
- **Monitoring and reporting:** Monitoring and reporting on the status of data processing pipelines, workflows and jobs
- **Compliance management:** Automating compliance tasks, such as data masking and archiving, to meet regulatory requirements

Hive is a data warehouse software project built on top of Apache Hadoop to enable data query and analysis. It has an SQL-like interface to query data stored in various databases and file systems that integrate with Hadoop. The capabilities include the following:

- **Data warehousing:** Storing and analysing large amounts of structured and semi-structured data
- **Customer analytics:** Collecting and analysing customer data to gain insights and improve customer experience
- **Fraud detection:** Detecting and preventing fraudulent activity through data analysis
- **Risk management:** Analysing large amounts of data to identify and manage financial risks
- **Loan processing:** Automating loan processing by analysing borrower data and reducing manual effort

Iceberg, a data management framework, offers the following capabilities to the banking sector:

- **Data warehousing:** Efficiently storing and querying large amounts of structured and semi-structured data
- **Customer analytics:** Collecting and analysing customer data to gain insights and improve customer experience
- **Fraud detection:** Detecting and preventing fraudulent activity through data analysis
- **Risk management:** Analysing large amounts of data to identify and manage financial risks
- **Loan processing:** Automating loan processing by analysing borrower data and reducing manual effort

Kafka is an open-source distributed event-streaming platform used by thousands of companies for high-performance data pipelines, streaming analytics, data integration and mission-critical applications. [3] Its capabilities cover the following:

- **Real-time data streaming:** Handling high volumes of real-time data such as financial transactions and sensor data
 - **Event-driven architecture:** Building event-driven systems for use cases such as fraud detection, risk management and customer analytics
 - **Message queue:** Providing a scalable and reliable message queue for distributing data between systems
 - **Data integration:** Integrating data from multiple systems into a centralised repository for analysis
 - **Compliance management:** Automating compliance tasks, such as data masking and archiving, to meet regulatory requirements
-

Conclusion

Emerging trends in big data for the banking space indicate a positive outlook. A key development is the use of machine learning and artificial intelligence to automate and improve decision-making processes. Banks and financial institutions are also exploring the use of blockchain technology to improve data security and data sharing.

The use of big data in banking is a key enabler for extracting meaningful insights from the vast amount of data generated and stored. Through advanced analytical techniques, banks can improve their capabilities in fraud detection, risk management, customer segmentation and personalised marketing. However, they face unique challenges in managing and analysing their data and must comply with strict regulations on data privacy and security. As big-data analytics continue to evolve, banks should keep up with the latest trends and best practices to remain competitive.

How Acuity Knowledge Partners can help

We play a vital role in assisting banking clients with their big-data needs. Our team of data engineers and development operations engineers helps clients manage and analyse their data effectively. They help automate and optimise data pipelines, develop and implement machine learning models for fraud detection and risk management, and provide real-time data analysis for improved decision-making. They also help ensure compliance with regulations around data privacy and security. Our expertise in big-data technologies such as Hadoop, Spark, Airflow, Hive and Kafka helps clients extract meaningful insights from their data and stay competitive.

References:

[1] Big data in banking:

https://www.researchgate.net/publication/361227210_The_role_of_big_data_in_financial_sector_A_rev

[2] Big-data application in banking sector:

<https://www.irjet.net/archives/V7/i6/IRJET-V7I61197.pdf>

[3] Impact of big-data analytics on the banking sector:

<https://doi.org/10.13140/RG.2.1.1138.4163>

[4] Toolkits for big-data processing:

About Acuity Knowledge Partners

Acuity Knowledge Partners is a leading provider of high-value research, analytics and business intelligence to the financial services sector. The company supports over 650+ financial institutions and consulting companies through a team of over 6,000+ subject matter experts who work as an extension of the clients' teams based out of various global delivery centres.

We empower our clients to drive revenues higher. We innovate using our proprietary technology and automation solutions. We enable our clients to transform their operating model and cost base.