KOGOD SCHOOL of BUSINESS
AMERICAN UNIVERSITY • WASHINGTON, DC

ITEC670: Databases and Big Data Research Paper:

Modern Database Management in Banking: Evolution, Scalability, and AI-Driven Fraud Detection

Yenjo Lee (Sally)

Last Updated: 28 April 2025

Introduction:

In recent decades, the banking industry has undergone a significant transformation in how data is stored, processed, and leveraged. From traditional on-premises relational databases to cloud-native distributed systems, financial institutions now rely on robust data architectures to support real-time decision-making, risk management, and regulatory compliance. As banking continues to digitize, modern database management systems (DBMS) are expected to handle vast volumes of structured and unstructured data while ensuring security, scalability, and operational efficiency (Ionescu et al., 2025).

My interest in this topic stems from my previous experience working at a bank, where I witnessed firsthand the importance of reliable, well-structured data in supporting day-to-day operations. During my graduate studies in data science, I've focused on financial analytics, and this paper builds upon that foundation by exploring the evolution of database technologies in the banking sector. I will also introduce a concept I've been developing a generative AI-powered fraud detection chatbot. While the full implementation is still in progress, I include the design approach and interface development to illustrate how modern data infrastructure can support such intelligent systems.

This paper thus aims to examine the role of database management systems in modern banking, provide an overview of machine learning techniques used in fraud detection, and briefly explore how these tools might integrate with generative AI to enhance real-time financial decision-making and user-facing applications.

Databases and Database Management Systems (DBMS)

A database is an organized collection of data that can be easily accessed, managed, and updated. In the context of banking, databases store crucial information such as customer profiles, account balances, transaction histories, credit scores, and fraud alerts. A Database Management System (DBMS) is a software that enables users and applications to interact with databases efficiently and securely. Key responsibilities of a DBMS include data storage, retrieval, modification, access control, and backup operations (Elmasri & Navathe, 2021).

Relational Database Management Systems (RDBMS), such as Oracle, Microsoft SQL Server, and MySQL, have traditionally dominated the banking sector. These systems use structured query language (SQL) and relational models to manage structured data with strict integrity constraints, which are ideal for financial records requiring high consistency and accuracy (Coronel & Morris, 2020).

Emerging Alternatives: NoSQL and NewSQL

As data volume and variety have increased, driven by mobile banking, digital wallets, and online transactions, banks have explored alternatives like NoSQL databases (e.g., MongoDB, Cassandra) and NewSQL platforms (e.g., Google Spanner, CockroachDB). NoSQL databases offer flexibility for semi-structured and unstructured data, scaling horizontally to support massive distributed environments (Margara et al., 2023, Schukin et al., 2024). This adaptability enables banks to capture diverse data types such as behavioral logs, transaction metadata, and fraud-related signals that do not conform to traditional relational schemas.

NewSQL databases offer the best of both worlds by combining the flexibility and scalability of NoSQL systems with the strong consistency and transactional reliability of traditional relational databases. This makes them a good choice for modern banking applications that need to process a large number of transactions quickly while maintaining data accuracy. Banks are increasingly using NewSQL solutions to support real-time analytics, fraud detection, and customer engagement (Maddali, 2021).

Why Data Matters in Banking

Data plays a key role in every part of banking today. Banks use data to manage customer accounts, track transactions, detect fraud, and meet government regulations. Good data management helps banks make faster and better decisions, such as offering the right financial products to customers or quickly spotting suspicious activities. It also helps banks avoid mistakes, stay in line with rules like privacy laws, and protect customer trust. Without reliable data, banks would struggle to operate safely and efficiently in a digital world where millions of transactions happen every day.

When managed effectively, data enables banks to:

- Detect and mitigate fraudulent transactions.
- Personalize financial products for customer needs.
- Ensure transparency and accuracy in compliance reporting.
- Increase operational efficiency and reduce processing costs.

Modern Use Cases in Fraud Detection

In recent years, banks have increasingly turned to machine learning (ML) to improve fraud detection. Compared to traditional rule-based systems, ML models can learn from past transaction data and identify hidden patterns that may signal fraud. Common methods include decision trees, logistic regression, and gradient boosting, which are effective for classifying whether a transaction is likely to be fraudulent. More advanced models like neural networks and deep learning can handle more complex data and are especially useful when fraud behavior changes quickly over time (Aziz et al., 2023). These models are often used in real-time systems that flag suspicious activity as soon as it happens.

In addition to numerical transaction data, natural language processing (NLP) techniques are now being used to analyze text data such as customer messages, dispute reports, or transaction notes. Large language models (LLMs), such as GPT, can help detect fraud by analyzing the language customers use and identifying unusual or risky behavior. These models are also being tested in interactive fraud chatbots that explain alerts or provide advice to users in real time. By combining structured transaction data with insights from text, banks can create more accurate and responsive fraud detection systems (Aziz et al., 2023).

Data Challenges and the Rise of Generative AI in Fraud Prevention

One major issue is class imbalance, fraudulent transactions make up only a tiny portion of all data, which can make it harder for models to learn effectively. To address this, techniques like oversampling or creating synthetic fraud cases are often used. Banks must also be careful with data privacy and security, especially when handling sensitive customer information. Another challenge is that fraud tactics constantly evolve, so models need to be updated frequently to stay effective. Ensuring fairness and transparency is also important, as biased or unexplainable models can lead to incorrect decisions and loss of trust (Chen et al., 2025).

At the same time, generative AI is emerging as a promising new tool in fraud prevention. Large language models (LLMs), which are often used in natural language processing, can now help detect fraud not only from transaction data but also from written communication like customer messages or support logs. These models can summarize suspicious behavior or even act as chatbots that explain fraud alerts in real time.

The Role of Structured Data in AI Systems

Structured data is very important for making AI systems work well in banking. Things like customer transaction records, account balances, and credit scores give AI models real information to learn from. Without this organized data, even the most advanced models would struggle to make accurate predictions or provide useful advice. In financial applications, structured data helps AI models better understand patterns, detect risks like fraud, and suggest products or services to customers (Yuqi et al., 2024).

Recently, large language models (LLMs) have started to work closely with structured databases to improve banking services. Instead of just generating general responses, LLMs can now pull real numbers, transaction details, and customer history from databases to give personalized and accurate answers. This combination makes financial AI systems much more reliable and practical for real-world tasks like fraud detection, customer support, and financial planning (Yuqi et al., 2024).

Design: Generative AI Fraud Detection Chatbot

To explore the intersection of database management and artificial intelligence, I designed and implemented a prototype for a Generative AI-powered Fraud Detection Chatbot tailored for banking applications. The system demonstrates how structured data storage, secure user authentication, and dynamic AI prompt engineering can be integrated to deliver personalized fraud risk assessments.

The chatbot architecture consists of several key components:

1. Database Layer

A structured user and transaction database underpins the system. The database stores user profiles, encrypted passwords, transaction histories, and indicators of potential fraud (e.g., foreign transactions, unusually large transactions). In the current prototype, this data is simulated using Pandas DataFrames, but the structure is compatible with migration to a relational database such as PostgreSQL.

2. User Authentication Module

Before accessing chatbot services, users must log in by providing their username and password. The system verifies credentials against stored records to ensure authorized access to sensitive transaction data, simulating real-world banking security practices.

3. Data Retrieval Module

Upon successful authentication, the system retrieves the authenticated user's recent transaction records and profile information. This data serves as the context for personalizing the AI's fraud risk evaluation and recommendations.

4. Prompt Engineering and AI Layer

When a user submits a query (e.g., "Am I at risk of fraud?"), the chatbot dynamically constructs a structured prompt that includes a summary of the user's recent transactions and the user's specific question. This prompt is sent to a generative AI model (Gemini-1.5-Pro) via API.
For example:

"User johndoe recently made the following transactions: $500 at USA on 2025-04-01, $4000 at France on 2025-04-02. The user asks: 'Am I at risk of fraud?'. Please evaluate the fraud risk and advise."

5. Response Generation

The AI model generates a natural language response, providing tailored advice based on the retrieved transaction patterns. The chatbot may recommend specific actions, such as changing passwords, enabling two-factor authentication, or monitoring account activity, depending on the risk detected.

6. Conversation Logging

All user inputs and AI responses are logged in to the system's memory during the session to enable multi-turn conversations. This logging allows the chatbot to maintain context across multiple user questions within a session and provides a foundation for future system auditing or fine-tuning.

This design demonstrates how effective database management and real-time AI processing can enhance fraud detection in banking systems, with flexibility for future extensions such as cloud database integration and advanced risk scoring.

Benefits and Challenges

The evolution of database management, combined with advances in artificial intelligence, offers banks significant operational and strategic advantages. Modern cloud-based databases enable financial institutions to dynamically scale their storage and processing power, accommodating growing transaction volumes and shifting customer demands. Distributed databases such as Google Spanner and Amazon Aurora support high availability and global access, critical for banks operating in multiple regions (Margara et al., 2023).

The integration of machine learning models into real-time data pipelines enhances the detection of sophisticated fraud patterns that traditional rule-based systems often miss. AI-driven fraud prevention tools, including intelligent chatbots and anomaly detection engines, can identify threats quickly and reduce false positives, thereby improving risk management outcomes (Chen et al., 2025). In addition, these systems significantly improve the customer experience. By leveraging accurate backend data, AI-powered services can offer personalized alerts, fraud warnings, and financial advice, helping to build trust and long-term customer loyalty (Aziz et al., 2023).

Moreover, automating tasks like fraud monitoring, compliance reporting, and answering customer questions helps banks work much more efficiently. Instead of relying only on human employees to manually check for fraud or complete long reports, AI systems and smart databases can handle many of these tasks automatically and in real time. This saves banks a lot of time and money, reduces human error, and allows employees to focus on more important work, such as improving customer service, creating new financial products, or making strategic business decisions (Gudipati, 2025).

However, there are also important challenges that come with using modern databases and AI in banking. Moving sensitive customer data to cloud systems increases the risk of cyberattacks and data breaches, making strong security a top priority. AI models can also introduce bias if they learn unfair patterns from historical data, leading to inaccurate or unfair decisions. Upgrading from old banking systems to new cloud-based or AI-driven platforms is complicated and costly, and if not handled carefully, it can cause service disruptions. Real-time fraud detection systems, while powerful, also require a lot of computing resources and can become expensive to maintain. Finally, many advanced AI models are difficult to fully explain, and banks must find ways to make their systems more transparent to meet growing regulatory requirements (Aziz et al., 2023).

Conclusion

Database management from traditional relational databases to modern distributed and cloud-native systems, banks have adapted their infrastructures to meet rising demands for scalability, flexibility, and security. These advancements not only support core banking operations but also enable the integration of emerging technologies such as machine learning and generative AI into fraud detection and customer service.

Machine learning techniques have become essential for identifying sophisticated fraud patterns that traditional rule-based systems often fail to detect. Real-time data processing architectures, combined with predictive modeling, empower banks to respond proactively to threats while enhancing customer trust and regulatory compliance. However, these innovations also bring challenges, including data security risks, model bias, integration complexity, and the need for explainability in AI-driven decision-making.

This paper also introduced the conceptual design of a generative AI-powered fraud detection chatbot, illustrating how modern database systems can support intelligent, personalized fraud prevention in future banking applications. Although the chatbot's full implementation is still in progress, the design highlights the potential for combining structured data management with real-time AI interaction to improve fraud detection and user engagement.

As banking continues to evolve in the era of big data and artificial intelligence, institutions must strike a careful balance between innovation and risk management. Investing in transparent, secure, and scalable database architectures, alongside responsible AI practices, will be key to building resilient, customer-centric, and future-ready financial systems.

References

Aziz, L. A. R., & Andriansyah, Y. (2023). The role artificial intelligence in modern banking: an exploration of AI-driven approaches for enhanced fraud prevention, risk management, and regulatory compliance. *Reviews of Contemporary Business Analytics*, *6*(1), 110-132.

Chen, Y., Zhao, C., Xu, Y., & Nie, C. (2025). Year-over-Year Developments in Financial Fraud Detection via Deep Learning: A Systematic Literature review. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2502.00201

Coronel, C., & Morris, S. (2020). *Database Systems: Design, Implementation, & Management* (13th ed.). Cengage Learning.

Elmasri, R., & Navathe, S. B. (2021). *Fundamentals of Database Systems* (7th ed.). Pearson.

Gudipati, S. S. (2025). AI-DRIVEN RISK MANAGEMENT AND FRAUD DETECTION IN FINANCIAL SERVICES: A TECHNICAL DEEP DIVE. *INTERNATIONAL JOURNAL OF INFORMATION TECHNOLOGY AND MANAGEMENT INFORMATION SYSTEMS*, *16*(1), 860–875. https://doi.org/10.34218/ijitmis_16_01_061

Ionescu, S., Diaconita, V., & Radu, A. (2025). Engineering sustainable data architectures for modern financial institutions. *Electronics*, *14*(8), 1650. https://doi.org/10.3390/electronics14081650

Margara, A., Cugola, G., Felicioni, N., & Cilloni, S. (2023). A model and survey of Distributed Data-Intensive Systems. *ACM Computing Surveys*, *56*(1), 1–69. https://doi.org/10.1145/3604801

Maddali, R. (2021). Machine Learning for SQL-Based Anomaly Detection & Fraud Analytics in Financial Data. *International Journal of Leading Research Publication.*, *2*(10). https://doi.org/10.70528/ijlrp.v2.i10.1456

Nie, Y., Kong, Y., Dong, X., Mulvey, J. M., Poor, H. V., Wen, Q., & Zohren, S. (2024). A survey of large language models for financial applications: Progress, prospects and challenges. arXiv preprint arXiv:2406.11903.

Schukin, A., Scerbakov, N., & Rezedinova, E. (2024). A data model for semi-structured data. *AIP Conference Proceedings*, *3001*, 070009. https://doi.org/10.1063/5.0200874

Zhang, Z., Megargel, A., & Jiang, L. (2025). Performance evaluation of NewSQL databases in a distributed architecture. *IEEE Access*, 1. https://doi.org/10.1109/access.2025.3529740