The North African Journal of Scientific Publishing (NAJSP)
مجلة شمال إفريقيا للنشر العلمي (NAJSP)
E-ISSN: 2959-4820
Volume 3, Issue 3, 2025
Page No: 243-250
Website: https://najsp.com/index.php/home/index
SJIFactor 2024: 5.49 | معامل التأثير العربي (AIF) 2024: 0.71 | ISI 2024: 0.696

A Comparative Study of Apache Flink and Spark in Real-Time Financial Data Analytics

Taghrid Z. Ali*
Software Development Department, Higher Institute of Engineering Technologies-Benghazi
Benghazi, Libya

دراسة مقارنة بين Apache Flink وSpark في تحليلات البيانات المالية في الوقت الفعلي

تغريد زيدان علي*
قسم تطوير البرمجيات، المعهد العالي للتقنيات الهندسية – بنغازي، بنغازي ، ليبيا

*Corresponding author: taghreed_zidan@hotmail.com

Received: July 14, 2025	Accepted: August 20, 2025	Published: August 24, 2025

Abstract:
With the rapid development of banking and financial technology, smart banking applications are generating ever-increasing data, making big data analysis in the banking sector no longer a luxury but a necessity for operating effectively in this data-driven digital environment. Data analytics in the banking sector can help generate in-depth insights into customer behavior, enhancing their ability to provide more personalized and proactive customer service, improving customer experience, and strengthening trust between them and their banking institutions. This study aims to shed light on the use of Internet of Things (IoT) technology in the banking sector and explore appropriate methods for analyzing the resulting data by it. In this context, the study compared the Apache Flink and Apache Spark platforms to determine the most appropriate solution for real-time data analysis tasks. This study contributes to bridging the existing gap in helping banking institutions choose the appropriate data analysis platform according to their needs. This comprehensive systematic review constitutes the basic methodology used by the author to investigate issues related to the integration of IoT technologies and data analytics within the emerging banking industry.

Keywords: Data analytics, Apache Flink, Apache Spark, Internet of Things.

الملخص
مع التطور السريع للتكنولوجيا المصرفية والمالية، تُنتج تطبيقات الخدمات المصرفية الذكية بيانات متزايدة باستمرار، مما يجعل تحليل البيانات الضخمة في القطاع المصرفي ليس مجرد رفاهية، بل ضرورة للعمل بفعالية في هذه البيئة الرقمية القائمة على البيانات. يمكن لتحليلات البيانات في القطاع المصرفي أن تُسهم في تكوين رؤى معمقة حول سلوك العملاء، مما يُعزز قدرتهم على تقديم خدمة عملاء أكثر تخصيصًا واستباقية، ويُحسّن تجربة العملاء، ويعزز الثقة بينهم وبين مؤسساتهم المصرفية. تهدف هذه الدراسة إلى تسليط الضوء على استخدام تقنية إنترنت الأشياء (*IoT*) في القطاع المصرفي، واستكشاف الأساليب المناسبة لتحليل البيانات الناتجة عنها. في هذا السياق، قارنت الدراسة بين منصتي *Apache Flink* و*Apache Spark* لتحديد الحل الأنسب لمهام تحليل البيانات في الوقت الفعلي. تُساهم هذه الدراسة في سد الفجوة القائمة في مساعدة المؤسسات المصرفية على اختيار منصة تحليل البيانات المناسبة وفقًا لاحتياجاتها. تُشكل هذه المراجعة المنهجية الشاملة المنهجية الأساسية التي استخدمها المؤلف لدراسة القضايا المتعلقة بدمج تقنيات إنترنت الأشياء وتحليلات البيانات في القطاع المصرفي الناشئ.

الكلمات المفتاحية: تحليلات البيانات، *Apache Flink*، *Apache Spark*، إنترنت الأشياء.

Introduction

Internet of things technologies has become a prominent feature of our time, as its technologies have diversified and overlapped in various fields of economic, medical, agricultural, and other life, which prompted institutions to move forward to keep pace with this acceleration. It is worth noting that, recent research indicates that the IoT has been identified as one of the most important and fastest-rising trends [1]. The IoT is expected to grow in the global economy, from $2.7 trln to $6.2 trln annually until 2025, and will be worth $14 trln by 2030 [2]. By 2025, there will be 25 billion IoT connections worldwide, To adapt to these changes, traditional banks must reconsider their business models and offerings [3].

Although many literature has addressed several different definitions of IoT: IoT is a method for exchanging information using the Internet to combine the physical world with the virtual world [4]. The Internet of Things (IoT) is an essential component of smart infrastructure and is critical to modern banking operations. In the banking sector, the term "IoT" refers to the integration of interconnected systems, sensors, and devices connected to the internet that collect, analyze, and exchange data to perform banking operations. This aims to create a more secure, personalized, and efficient financial environment, with the goal of improving and developing the performance of banking operations [5] [6].

Banks collect customer data from Through Biometric sensors and smart mobile apps for use within IoT technologies. In this regard, Smart apps are also an IoT technology, through which banking services are provided. In fact, Most of these apps miss the advantage of data analysis, so banks do not benefit from an important and reliable source of data. Knowing what the customer wants, identifying their needs, and linking them to a specific marketing activity enables banking institutions to achieve great profits. Furthermore, Maintaining existing customers and attracting more new customers depends on the presence of marketing activities supported by electronic banking services, and this is what the banking sector needs. As already mentioned, The use of Internet of Things (IoT) technologies is linked to the continuous increase in data volume. Despite all the advantages associated with using IoT technologies in the banking sector, they face challenges related to managing and analyzing massive amounts of data in real time. This calls on banks to establish their infrastructure to address this continuous increase in data volume. Many banking institutions in Libya have provided digital banking services, It can be said that these emerging applications are the basic building block of IoT technologies. To keep pace with rapid technological progress, the banking sector must develop the technologies used to respond to customers' requirements.

In this research paper, we aim to highlight the Internet of Things (IoT) technology within the banking sector and link it to the data analysis process, specifically in cases that require real-time data processing. To achieve this, we will conduct a comparison between the two frameworks (Apache Flink and Apache Spark) based on their key performance indicators, in order to measure the performance of each platform, determine its effectiveness in data analysis, and evaluate its impact on the efficiency of banking operations. Then, the most appropriate framework will be determined for analyzing banking data resulting from the use of IoT applications in real-time. The information derived from this comparison makes it easier to identify best practices for selecting frameworks based on the specific application needs of banking institutions. The rest of this paper is organized into following five sections: reviews related work in section II. Section III describes the IOT in banking services . The data analytics is described in Section IV, whereas Section V discusses the key performance criteria for comparing. Research discussion and results are then presented in section VI. Conclusion is discussed in section VII.

Related Works

Recent scholarly inquiries have substantially advanced the understanding of the intricate challenges linked to data analysis that stem from the implementation of Internet of Things (IoT) technology within the banking industry across various contexts. Each scholarly inquiry enriches the understanding of the implications of data analysis and its capacity to enhance customer insights, promote informed decision-making, and improve the efficiency of banking operations.The subsequent studies offer critical perspectives on the incorporation of IoT technology into banking applications, emphasizing the data analysis dimension within this particular domain. Moreover, these studies underscore several areas that necessitate additional inquiry, thereby revealing a research void that merits further examination and exploration. The following section presents a review of the most salient of these investigations

The study in [7] highlights modern methods for data analysis and explains the importance of analyzing Internet of Things data in real-time and its integration with artificial intelligence algorithms.It also reviewed the benefits of using advanced analytical methods in banks. Nevertheless, the study did not provide specific platforms or tools to facilitate this integration. Moreover, the work described in [8] has investigated a conceptual framework that enables continuous banking audits through real-time auditing and analysis of data generated by continuously streaming Internet of Things devices. The design of the proposed framework is based on improving the audit process without detailing the technical infrastructure or software supporting data analysis processes.

The work conducted in [9] aimed to present a systematic literature review that demonstrated the reality of the Internet of Things (IoT) and big data within the field of electronic banking services and studied their impact on various banking sectors, such as risk analysis and management, credit services, and financial operations. The study also reviewed the challenges arising from integrating the Internet of Things (IoT) and big data into banks and provided solutions to address these challenges. The study highlighted various important issues related to data analysis, such as data privacy, security, and data quality, and their importance in analyzing big data and its impact on decision-making. However, the study did not provide specific technical solutions or indicate platforms for performing data analysis operations.

The study in [10] explains the importance of integrating Internet of Things (IoT) and Big Data (BCT) technologies, as this facilitates the process of collecting and processing data in real time. In this context, the study presented several use cases in which the Internet of Things (IoT) and Big Data were integrated in various fields, such as digital banking and secure banking services. On the other hand, the study did not clarify the technologies required for processing or analyzing data. Another study conducted by [11] explains the impact of big data analytics on financial operations, such as monitoring customer feedback and its role in financial and banking services. Regarding data analytics techniques, the study did not explicitly address this in the contexts presented. Instead, it focused on explaining the effects and benefits of using big data analytics .

The study described in [12] highlighted Internet of Things technology and the use of the resulting data to extract customer insights in the banking sector. In this context, the study proposed an architecture specifically designed for the banking sector that supports techniques for managing and analyzing data generated by various devices related to customer usage. Through the previous review of studies related to this research area, we noted that the research that indicated the presence of Internet of Things technology within the technical infrastructure of the banking sector in various contexts did not discuss the technologies that can be used for data analysis, but rather merely referred to them in passing. Therefore, author seek to bridge this gap by proposing an appropriate solution to create a banking technical infrastructure that is appropriate and supportive of Internet of Things data analysis within the banking sector.

IoT in Banking Services

As customers increasingly embrace banking applications and prefer digital services, traditional banks are forced to intensify their efforts toward deploying smart applications that support their operating models to maintain their position in the market. In this section, we delve into the Internet of Things (IoT) technologies used today in banking services, which are shaping the digital future of the banking sector.

Today's consumers are looking for more convenient and efficient ways to conduct banking transactions. In this regard, smart ATMs, interactive self-service kiosks, and wearable devices offer a wide range of benefits to financial institutions. Smart ATMs offer services such as real-time monitoring of banking transactions and interactive self-service for customers, enabling them to complete tasks such as paying bills, depositing checks, and managing accounts without human intervention. Therefore, IoT technologies support smart ATMs, enabling them to detect various fraud patterns and contribute to real-time prevention. Customer service platforms and wearable devices also enable customer service, enabling them to conduct banking transactions without using their physical cards [13].

Point of Sale (POS) terminals enhance customer service. They are an interactive interface consisting of hardware and software that enables the merchant or service provider and the customer to conduct and manage financial transactions digitally and securely. POS systems are designed to be connected to the network to process customer transactions in real time, access customer data, and manage inventory. POS devices also offer multiple payment methods, such as cash, credit cards, and mobile payments [14]. POS functionality has been integrated into smartphone applications, allowing customers to conduct transactions directly from their smartphones. Integrating IoT technology with POS systems enables banking institutions to enhance customer service and improve their shopping strategies. By tracking purchases and identifying customer preferences, they can target customers with promotional offers to market products more relevant to their needs. Recognizing this necessity, banks must prioritize digital initiatives that combine the use of smart mobile banking services with IoT technologies. This strategic focus is not just a response to current trends in digital banking transformation, but also a proactive effort to secure the future of its business in a smartphone-dependent world.

Data Analytics (DA)

Data Analytics (DA) are techniques and tools for analyzing large and small data sets with diverse data attributes to obtain meaningful conclusions and actionable insights [15]. The conclusions come in a variety of forms such as patterns and statistics, which enable organizations to deal with data more proactively in order to implement more effective decision-making processes [16]. Big data is processed and analyzed to extract meaningful insights and insights, which traditional methods are unable to deliver adequately. Working on analyzing the data provided by IoT technologies enables institutions to extract customer insights and thus create many opportunities for banking services [17]. It enables real-time analysis of customer behavior by integrating IoT technology, artificial intelligence, and customer behavior prediction, which enables banks to prevent fraud and monitor and track banking assets to ensure their security [7]. The implementation of the IoT becomes more efficient if the data analytics process is integrated with the IoT architecture [12].

Key Performance Criteria For Comparing

This section presents and discusses the most popular platforms for big data analytics by reviewing their associated characteristics. Apache Spark and Apache Flink, both frameworks provided by the Apache Software Foundation, were selected as the most widely used and have been reviewed by numerous researchers. We can compare the two frameworks based on their key performance indicators. This comparison aims to highlight the features, issues, and benefits of each, as well as their associated advantages. The insights gained from this comparison help identify and select the most appropriate framework based on the needs of the organization seeking to analyze its data .If we want to evaluate the efficiency of both Apache Flink and Apache Spark's performance in processing and analyzing customer transaction data in banking applications, we must set out a set of key performance criteria for each. The criteria were defined based on an analysis of the perspectives and insights drawn from the informed research contexts.

Table 1: Comparison of Key Performance Indicators between Apache Flink and Apache Spark.

Platform	Key
-	Performance and Throughput
Apache Flink	Low latency, high-throughput processing and stateful computations[18].
Apache Spark	It can achieve High throughput, with micro-batch architecture but at the cost of higher latency[18].
	Processing Models
Apache Flink	optimized for efficiently handle stream and batches processing[19].
Apache Spark	Excels in batch processing, and supports stream processing through micro-batch and structured streaming [20].
-	Scalability in Real-Time Data Processing
Apache Flink	Flink's architecture supports linear scalability, enhancing its scalability and reliability in data ingestion[21].
Apache Spark	It achieves scalability through its distributed computing model, which allows it to process data across multiple nodes in a cluster [22]
	Fault tolerance
Apache Flink	- A checkpointing mechanism that enables processing of state flows and aids in recovery without data loss. - Use of a circular cache to improve performance in high-throughput scenarios. - Stability and fast recovery times[23].
Apache Spark	Spark relies on lineage-based fault recovery, where lost data is reconstructed from previous computations, which can lead to increased event latency [24].
-	Security
Apache Flink	The fault-tolerance mechanisms include a distributed architecture, the use of checkpointing, and stateful processing, which helps in dealing with faults effectively. Its ability to handle continuous data streams also allows for early detection of faults or abnormal patterns [18].
Apache Spark	It uses several mechanisms to enhance security, such as fine-grained access control, data encryption, and authentication via protocols such as Kerberos. It also provides data integrity verification [25].
-	Machine learning
Apache Flink	Flink-ML library supports various algorithms specifically designed for real-time data processing, such as Classification Algorithms, Clustering Algorithms, Regression Algorithms, and Collaborative Filtering[26].
Apache Spark	Spark is based on the machine learning library MLLIB, which includes multiple types of algorithms such as: Gradient Boosting , Logistic Regression, Decision Trees ,Random Forest, and XGBoost [27] [28].
-	Memory Management

Apache Flink	Flink uses various advanced memory management mechanisms. It leverages CPU-GPU integration to improve load balancing and increase computational capacity through efficient task segmentation and scheduling. It also uses Task Requirement-Based Allocation to allocate memory to suit task needs. Lifetime-Based Management analyzes data object lifetimes, which improves memory allocation and reduces garbage collection overhead [29-31] .
Apache Spark	Memory can be managed using various mechanisms: OOM Handling, Adaptive Memory Reservation , ATuMm Implementation, Data Compression and Serialization, Fine-Grained Monitoring. Although these approaches enable Spark to manage memory efficiently, challenges remain, particularly in balancing memory allocation needed to meet the real-time data processing needs of banking applications [32-35]
-	Use Cases
Apache Flink	Suitable for real-time transaction processing and stateless computations [36].
Apache Spark	Suitable for hybrid workloads, including batch and machine learning tasks [37].
	Resource Utilization
Apache Flink	Real-time processing at a cost-effective level[38].
Apache Spark	Batch tasks are more economical than stream processing due to higher costs[37].
-	Ecosystem
Apache Flink	An ecosystem that is expanding and has strong stream processing support [36].
Apache Spark	A mature ecosystem that has extensive libraries for machine learning and SQL [37].

Discussion And Results

The comparative metrics in Table 1provide a clear view of the strengths and weaknesses of Apache Spark and Apache Flink. In order to determine which framework better satisfies the requirements of real-time data analytics in IoT-enabled banking contexts, we will analyze these metrics in-depth in this section, drawing on scholarly literature already in existence. In contexts necessitating continuous streaming data processing, Flink attains reduced latency and augmented throughput as a result of its proficiency in managing real-time data with efficacy. Conversely, Spark exhibits a superior throughput capacity that can be ascribed to its micro-batch framework; nevertheless, this benefit is counterbalanced by an escalation in latency.Consequently, one could posit that Flink exhibits superior performance in the context of real-time data processing, whereas Spark is more appropriately aligned with batch-oriented operations. The Spark platform relies on micro-batch data processing, which in turn leads to latency. Therefore, Spark faces challenges with repetitive tasks and continuous data flow, requiring alternative solutions to improve efficiency. On the other hand, the Flink platform architecture enables it to scale more efficiently for operations that require real-time data processing and iterative processes, and it also contributes to improving the efficiency of iterative algorithms [39].

Apache Flink outperforms Apache Spark in tasks where real-time data processing is a critical factor, as Flink demonstrates highly efficient performance in terms of memory management and reducing the overhead caused by garbage collection. Furthermore, Apache Spark still has a number of limitations related to processing speed, the balance of memory usage, and the scalability when handling large-scale datasets. Flink offers several advantages when processing real-time data, the most important of which is fault tolerance, which includes faster recovery and lower latency time .The multiple security methods used in Spark make it a suitable choice for banking applications that require secure data sharing. Spark therefore outperforms the Flink platform in this regard, as Flink's architecture is primarily built for rapid, real-time processing.

The Flink-ML library supports a diverse set of efficient algorithms. It also speeds up model development because it's designed to simplify end-to-end machine learning pipelines and allows for immediate response to any data update due to the real-time nature of data processing. In contrast, the Spark platform supports machine learning via its MLlib library, which is designed to support both batch and micro-batch data processing. This makes the platform a strong contender for use in banking application data analysis. Banks pay considerable attention to their technology infrastructure due to its associated costs. Resource utilization is measured by the platform's ability to perform tasks with minimal resource consumption and operational costs.

In this direction, flink demonstrates greater operational efficiency than Spark when it comes to real-time data processing. Flink utilizes resources efficiently thanks to its streaming-native processing capability. It also manages memory efficiently, reducing memory consumption and associated operating costs. Flink also boasts horizontal scaling, which reduces the need for additional resources when processing large data streams. This reduces operational costs, resulting in scalability with lower costs. Flink also achieves lower latency compared to Spark Streaming, which means faster data analysis, which positively impacts decision-making in banking institutions and enhances business efficiency. In contrast, Spark consumes more resources due to its reliance on micro-batch processing, which leads to resource consumption and short idle periods. Production costs also increase in Spark because it often consumes more CPU and memory resources to process the continuous flow of data. The SPARK ecosystem is more mature than the Flink ecosystem, where we find it supports graphic processing and includes a wide range of machine learning libraries in return. Flink's ecosystem focuses mainly on stream processing [41] [40].

In light of the above, we can conclude that Apache Flink boasts several advantages, including fast processing and low latency, making it an effective solution for financial institutions seeking to make critical and rapid decisions, such as monitoring changes in financial markets or detecting fraudulent activities that require rapid intervention to address and prevent them in real time. Furthermore, Flink supports a dedicated machine learning algorithm library, enabling Flink to develop predictive models that enable banking institutions to conduct immediate AI-based analysis, helping them detect fraudulent patterns and identify customer preferences. On the other hand, compared to Spark, the Flink platform requires higher latency and does not provide immediate insights in situations requiring real-time data processing. Furthermore, the Flink platform is characterized by its great ability to scale horizontally. In cases requiring increased load handling, the distributed architecture of the Flink platform allows the addition of new nodes to the cluster, ensuring continued operation even in the event of a failure of some nodes. In contrast, the Spark platform requires increased resource utilization and operational overhead when processing real-time data using micro-batch technology when used in applications for large financial institutions. This results in increased maintenance costs and complexity within the Spark framework due to the use of different tools for batch and flow processing.

Conclusion

Through this study, we seek to improve the technological infrastructure in the banking sector, making it more effective, efficient, scalable, and secure .To achieve this, we worked through this study to identify the most appropriate platform for analyzing real-time data from Internet of Things devices. Therefore, choosing the appropriate data analysis platform depends on the specific needs of the targeted banking application. In conclusion, the research results confirm that Flink is most suitable for situations where real-time data processing is critical and necessary to generate immediate insights from the data .It is worth noting that this study was conducted on a limited number of performance indicators and was limited to only two types of data analysis platforms. Therefore, we suggest conducting further studies on other platforms and with different performance metrics. We also recommend directing future research to a real-world case study to ensure the applicability of the findings in a real-world setting.

References

[1] K. Ghaffari, M. Lagzian, M. Kazemi, and G. Malekzadeh, "A comprehensive framework for Internet of Things development: A grounded theory study of requirements," Journal of Enterprise Information Management, vol. 33, no. 1, pp. 23-50, 2020.

[2] M. S. Tokareva, K. O. Vishnevskiy, and L. P. Chikhun, "The impact of the Internet of Things technologies on economy," Бизнес-информатика, no. 3 (45) eng, pp. 62-78, 2018.

[3] F. Khanboubi, A. Boulmakoul, and M. Tabaa, "Impact of digital trends using IoT on banking processes," Procedia Computer Science, vol. 151, pp. 77-84, 2019.

[4] O. Elijah, T. A. Rahman, I. Orikumhi, C. Y. Leow, and M. N. Hindia, "An overview of Internet of Things (IoT) and data analytics in agriculture: Benefits and challenges," IEEE Internet of things Journal, vol. 5, no. 5, pp. 3758-3773, 2018.

[5] P. Thirumagal, T. Das, S. Das, E. Khatri, and S. Rani, "The Role of IoT in Revolutionizing Payment Systems and Digital Transactions in Finance," in 2024 5th International Conference on Recent Trends in Computer Science and Technology (ICRTCST), 2024: IEEE, pp. 171-176.

[6] D. Gupta and P. Kulkarni, "Internet of Things (IoT) driven digital transformation in BFSI sector," in 2023 Somaiya International Conference on Technology and Information Management (SICTIM), 2023: IEEE, pp. 55-58.

[7] P. Thirumagal, S. Vaddepalli, T. Das, S. Das, S. Madem, and P. Immaculate, "AI-Enhanced IoT Data Analytics for Risk Management in Banking Operations," in 2024 5th International Conference on Recent Trends in Computer Science and Technology (ICRTCST), 2024: IEEE, pp. 177-181.

[8] T. Yadav, P. Kulkarni, and A. Chirputkar, "Continuous Bank Audit Using Internet of Things," in 2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA), 2023: IEEE, pp. 1412-1417.

[9] P. Ebrahimi, D. K. Moghaddam, and Y. S. S. Mehrabani, "Challenges and opportunities of big data and IoT in the electronic banking industry: a systematic literature review," 2022.
[10] N. Surekha, R. Sangeetha, C. Aarthy, R. Kavitha, and R. Anuradha, "Leveraging blockchain technology for internet of things powered banking sector," in Blockchain based Internet of Things: Springer, 2022, pp. 181-207.
[11] C. Prabhu et al., "Big data analytics for financial services and banking," Big data analytics: Systems, algorithms, applications, pp. 249-256, 2019.
[12] V. Dineshreddy and G. Gangadharan, "Towards an “Internet of Things” framework for financial services sector," in 2016 3rd International Conference on Recent Advances in Information Technology (RAIT), 2016: IEEE, pp. 177-181.
[13] T.-H. Le, N. T. Dung, D. T. N. Huy, N. T. P. Thanh, D. T. N. Hien, and N. T. Hang, "Internet of Things (IOT) uses and applications-solutions in emerging markets and Vietnam," Turkish Journal of computer and mathematics education, vol. 12, no. 11, pp. 5546-5550, 2021.
[14] H. Ramalingam and V. P. Venkatesan, "Conceptual analysis of Internet of Things use cases in Banking domain," in TENCON 2019-2019 IEEE Region 10 Conference (TENCON), 2019: IEEE, pp. 2034-2039.
[15] A. Pavithra, C. Anandhakumar, and V. N. Meenashisundharam, "Internet of Things with BIG DATA Analytics–A Survey," International Journal of Scientific Research in Computer Science Applications and Management Studies, vol. 8, no. 1, pp. 2319-1953, 2019.
[16] Y. Perwej, M. A. AbouGhaly, B. Kerim, and H. A. M. Harb, "An extended review on internet of things (iot) and its promising applications," Commun. Appl. Electron, vol. 7, no. 26, pp. 8-22, 2019.
[17] N. Kumar et al., "Harnessing the power of big data: Challenges and opportunities in analytics," Tuijin Jishu/Journal of Propulsion Technology, vol. 44, no. 2, pp. 363-371, 2023.
[18] A. Sharma, D. Puri, M. Kumar, and G. Soni, "Implementation and comparison of big data analysis on large dataset using spark and flink," in ICCCE 2021: Proceedings of the 4th International Conference on Communications and Cyber Physical Engineering, 2022: Springer, pp. 385-394.
[19] L. K. Yenduri, "Performance Evaluation of Apache Hadoop, Spark, and Flink for Batch Processing of Big Data: A Comparative Analysis," in 2024 Third International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT), 2024: IEEE, pp. 1-5.
[20] Q. Le, M. Chen, and W. Chen, "Research on stream processing engine and benchmarking framework," in 2022 IEEE International Conference on Networking, Sensing and Control (ICNSC), 2022: IEEE, pp. 1-5.
[21] A. Bozkurt, F. Ekici, and H. Yetiskul, "Utilizing Flink and Kafka Technologies for Real-Time Data Processing: A Case Study," The Eurasia Proceedings of Science Technology Engineering and Mathematics, vol. 24, pp. 177-183, 2023.
[22] A. Ed-daoudy, K. Maalmi, and A. El Ouaazizi, "A scalable and real-time system for disease prediction using big data processing," Multimedia Tools and Applications, vol. 82, no. 20, pp. 30405-30434, 2023.
[23] Z. Zhang, W. Li, X. Qing, X. Liu, and H. Liu, "Research on optimal checkpointing-interval for flink stream processing applications," Mobile Networks and Applications, pp. 1-10, 2021.
[24] A. Vogel, S. Henning, E. Perez-Wohlfeil, O. Ertl, and R. Rabiser, "A comprehensive benchmarking analysis of fault recovery in stream processing frameworks," in Proceedings of the 18th ACM International Conference on Distributed and Event-based Systems, 2024, pp. 171-182.
[25] T. Xue et al., "GuardSpark++: Fine-grained purpose-aware access control for secure data sharing and analysis in Spark," in Proceedings of the 36th Annual Computer Security Applications Conference, 2020, pp. 582-596.
[26] M. Mezati and I. Aouria, "Flink-ML: machine learning in Apache Flink," Brazilian Journal of Technology, vol. 7, no. 4, pp. e74577-e74577, 2024.
[27] K. K. Pandey, A. Giri, S. Sharma, and A. Singh, "Predictive Analysis of Classification Algorithms on Banking Data," in 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON), 2021: IEEE, pp. 1-5.
[28] M. N. Hossain et al., "ENHANCED BANKING FRAUD DETECTION: A COMPARATIVE ANALYSIS OF SUPERVISED MACHINE LEARNING ALGORITHMS," American Research Index Library, pp. 23-35, 2024.

[29] W. Xiu, J. Guo, and Y. Li, "A memory management strategy based on task requirement for in-memory computing," in 2020 Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), 2020: IEEE, pp. 406-412.
[30] L. Lu et al., "Lifetime-based memory management for distributed data processing systems," arXiv preprint arXiv:1602.01959, 2016.
[31] C. Chen, K. Li, A. Ouyang, Z. Zeng, and K. Li, "GFlink: An in-memory computing architecture on heterogeneous CPU-GPU clusters for big data," IEEE Transactions on Parallel and Distributed Systems, vol. 29, no. 6, pp. 1275-1288, 2018.
[32] D. Jia, J. Bhimani, S. N. Nguyen, B. Sheng, and N. Mi, "Atumm: Auto-tuning memory manager in apache spark," in 2019 IEEE 38th International Performance Computing and Communications Conference (IPCCC), 2019: IEEE, pp. 1-8.
[33] D. Chen, H. Chen, Z. Jiang, and Y. Zhao, "An adaptive memory tuning strategy with high performance for Spark," International Journal of Big Data Intelligence, vol. 4, no. 4, pp. 276-286, 2017.
[34] W. Zhang and J. Li, "Spark Memory Management," in International Conference on Advanced Hybrid Information Processing, 2017: Springer, pp. 65-69.
[35] B. Li et al., "Adaptive memory reservation strategy for heavy workloads in the Spark environment," PeerJ Computer Science, vol. 10, p. e2460, 2024.
[36] O.-C. Marcu, A. Costan, G. Antoniu, and M. S. Pérez-Hernández, "Spark versus flink: Understanding performance in big data analytics frameworks," in 2016 IEEE International Conference on Cluster Computing (CLUSTER), 2016: IEEE, pp. 433-442.
[37] M. A. Lopez, A. G. P. Lobato, and O. C. M. Duarte, "A performance comparison of open-source stream processing platforms," in 2016 IEEE Global Communications Conference (GLOBECOM), 2016: IEEE, pp. 1-6.
[38] F. Ullah, S. Dhingra, X. Xia, and M. A. Babar, "Evaluation of distributed data processing frameworks in hybrid clouds," Journal of Network and Computer Applications, vol. 224, p. 103837, 2024.
[39] G. E. Gévay, T. Rabl, S. Breß, L. Madai-Tahy, J.-A. Quiané-Ruiz, and V. Markl, "Efficient control flow in dataflow systems: When ease-of-use meets high performance," in 2021 IEEE 37th International Conference on Data Engineering (ICDE), 2021: IEEE, pp. 1428-1439.
[40] K. Lang and X. Chai, "Implementation of load balancing algorithm based on flink cluster," in 2022 International Conference on Cloud Computing, Big Data Applications and Software Engineering (CBASE), 2022: IEEE, pp. 264-268.
[41] V. N. Edapurath, "Design and Implementation of a Scalable Distributed Machine Learning Infrastructure for Real-Time High-Frequency Financial Transactions," 2023.