Improve Real - Time Fraud Detection with DataOps on Resilient Elastic Platforms

Aparna Krishna Bhat

Senior Analyst, EY (Ernst & Young), USA

Abstract: ***Fraud detection refers to measures put in place to prevent criminals from obtaining monetary benefits through false claims. In the world of online commerce, scams, scams and malicious agents are harmful in many ways. Businesses should take steps to ensure that fraud is detected and stopped before it affects the business. Fraud prevention refers to the countermeasures in place to mitigate the impact that fraudsters can have on business operations, once detected. Fraud detection is the first step in identifying where the risk lies. Real - time fraud detection improves on - site management of fraudulent activities and channels that could otherwise lead to negative business outcomes. As fintech and e - commerce thrive, more and more bank payments and money transfers are facilitated through online channels, which are faster, more convenient and safer for health in the age of the coronavirus. Real - time fraud detection and prevention can be accomplished using fraud detection software, RiskOps tools, DataOps, and other risk management strategies that improve data usability. Fraud detection on elastic platforms like elastic search has the ability to detect information in real time through predefined standards and approaches that provide alerts when communication is imminent. However, the fraud detection approach works with a significant commitment to ensure consistent compliance with privacy rules, such as anonymization, which ensures that personally identifiable information is not used for malicious purposes. Therefore, the act of fraud detection requires instrumental data processing mechanisms and relies on the scalability and flexibility of elastic platforms to achieve a scalable operation.***

Keywords: AI, DataOps, RiskOps, Elasticsearch, Elastic Platforms, Fraud Detection, Compliance, Data management, Data pipeline, Data processing, Real time data processing, Data enrichment

1. Introduction

Fraud detection is a process that involves identifying fraudulent transactions in a system. The process also involves the use of different approaches to ensure the instrumental prevention of any fraudulent activity. Fraudulent activity comes in many forms, including identity theft, insurance fraud, financial fraud, and cybercrime. According to the Mastercard Global Consumer Study 2020, 8 out of 10 Mastercard users worldwide use contactless payment methods. The data on the transactions carried out are collected in the databases of banks, e - commerce platforms and other participants in the e - commerce sector [1]. Modern organizations must take action against fraud, improving fraud detection and working to ensure that outstanding issues in fraud management are resolved in a meaningful way. Fraud detection affects businesses in a variety of ways. These approaches allow the company to differentiate itself in terms of performance and reputation, attracting and guaranteeing sustainable added value at all levels. The proper use of data allows companies to improve operational performance and customer experience. However, given the large volume of data, it is impossible for organizations to manage it manually. Therefore, companies are forced to build a big data infrastructure and use machine learning (ML) algorithms. Developing mechanisms to protect customer funds from fraudsters is part of the strategy of banks and e - commerce companies to improve the customer experience. As payments move to online channels, fraudsters have also adapted, highlighting the problems with protecting funds that pass through online channels. According to SmartMetric, in 2018, global losses due to payment fraud exceeded $24 billion. Despite the difficulties in fraud detection (caused by fraudsters imitating normal customer behavior and the use of social engineering techniques), the fraud detection systems of large companies can detect and prevent up to 98 % of cases [2]. In terms of ML metrics, this means achieving a high recall rate. Today, the problem of a large number of false positives (FP) arises, which causes the problem of low accuracy or high false positive rate (FPR). On average, only one in five fraudulent transactions was blocked and one in six users were blocked by mistake during the year [4]. False positives lead to financial losses for companies investigating cases and contacting the customer, as well as increased pressure on the call center, as well as lost revenue due to transaction rejections. Considering one of the possible response strategies - calling the customer - the cost of managing a blocked transaction varies from 1.5 euros (depending on the bank that provided us with the data for our research) to 5 euros (European Central Bank) [5]. In addition, the more demanding customers become about the quality of services, the more false positives damage the company's reputation and reduce customer loyalty.

2. Fraud Detection Advantages

Some of the main advantages of fraud detection in companies are:

1) **Maintain Trust and Reputation:** When companies face fraud and suffer losses, their customers and stakeholders develop less trust and confidence in their ability to manage business challenges. Management is expected to maintain safe practices, and their inability to do so demotivates them to make a compelling call to management and achieve a credible outcome. However, detecting fraud in the business and avoiding challenges illustrates a level of credibility for the business, fostering greater trust in the business.
2) **Improve Customer Satisfaction:** Businesses thrive on happy customers and keeping their data secure is essential for them as well as for organizations. Providing high - level security and privacy as a powerful payment protection solution shows customers that you are a company they can trust, increasing their overall

DOI: https://dx.doi.org/10.21275/SR241001074626

satisfaction and trust. When customers feel that their information is safe, they are more likely to stay loyal to you, do business with you again, and recommend you through positive word of mouth, helping your business grow.

3) **Protection of Sensitive Information:** Fraud detection helps companies protect sensitive information for both the organization and its stakeholders. Confidentiality in the company and the right to privacy must be respected, ensuring the educational requirement to understand and address the important requirements to obtain suitable results for the company [6].
4) **Save Money: -** Payment fraud affects thousands of businesses each year, with losses from e - commerce fraud alone estimated at $48 billion by 2023. By implementing effective fraud prevention strategies, businesses can protect yourself from the risks of unauthorized transactions. and account taking, maintaining financial integrity and healthy profit margins. Additionally, an effective fraud prevention solution means less time and money spent investigating and resolving incidents, resulting in significant savings and smoother operations.
5) **Compliance with Regulatory Requirements:** The fraud detection process aims to fulfill the elements necessary to manage regulatory requirements. Regulatory management and compliance help the company avoid sanctions and fines, thus improving its ability to make legal transactions and maintain business continuity [9].
6) **Improve Decision - Making:** Accurate and reliable financial data is essential to make informed decisions about your business. With reliable information at their fingertips, businesses can plan more effectively, identify growth opportunities and navigate the market with confidence. This leads to smarter strategies and gives them a competitive edge.

Therefore, fraud detection is shown to be an important element of organizational management [7]. Fraud detection uses the ability of the business to manage its needs, derived from the need to meet and mark the progressive requirements in all approaches [8]. Therefore, fraud detection, thanks to its extraordinary advantages, promotes the growth and advancement of the business to obtain the best results in all aspects and analysis. The principle of automation aims to ensure that repetitive processes are performed automatically to improve efficiency and reduce the possibility of an error [10].

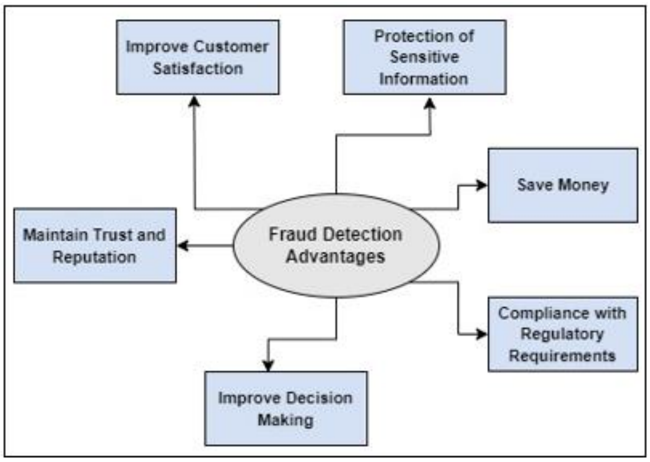
Improve Customer Satisfaction
Protection of Sensitive Information
Save Money
Maintain Trust and Reputation
Fraud Detection Advantages
Compliance with Regulatory Requirements
Improve Decision Making

Figure 1: Fraud Detection Advantages

3. DataOps and Elastic Data Platforms

DataOps is a collection of data management practices that aim to optimize delivery, quality control, collaboration, and data value maximization. These processes are managed by a set of principles that seek to enhance and appeal to various requirements in achieving and detailing a sustainable way of looking into the management and handling of various needs. The primary goal of DataOps is to break open silos between data producers (upstream users) and data consumers (downstream users) ensuring the access to reliable data sources. Thus, DataOps are a set of activities which enhance collaboration, automation and integration of data processes within the data life cycle. These processes are managed by a set of principles that seek to enhance and appeal to various requirements in achieving and detailing a sustainable way of looking into the management and handling of various needs.

The first collaboration principle aims to improve the cross - functional engagement of different experts, from business parties to data scientists and engineers. The second principle of continuous integration and delivery (CI/CD) involves rapid development and testing, as well as the distribution of data and analytics. The principle of automation aims to ensure that repetitive processes are performed automatically to improve efficiency and reduce the risk of error [10]. However, the monitoring and feedback aspect aims to evaluate and manage real - time analytics and ensure that feedback on each process is provided for the continuous improvement of the system. Elastic Slack is a comprehensive ecosystem of open - source tools for data ingestion, enrichment, storage, analysis and visualization. In addition to Elastic search, other software includes Logstash, Kibana, and Beats [11]. Each of these components have the capacity to ensure scalable and flexible avenue for the analysis, search and storage of huge datasets in real time. They can be used in various instances to enhance on the possibility and capacity to achieve remarkable development of values. For instance, they are used for security information and event management activities, enhancing the detection and response on security threats in the network in real time [11]. They allow for log and event analysis, ensuring troubleshooting and monitoring IT infrastructure across various application bases. They also apply in the capacity to observe data, enabling organizations to address the performance and status of their systems and applications across an instrumental scope of management. A major benefit of this system is the capacity to enable wide range of uses, ensuring companies can address their needs in various ways. The open - source nature of the platforms also provides a reliable scope of addressing organizational needs at all times.

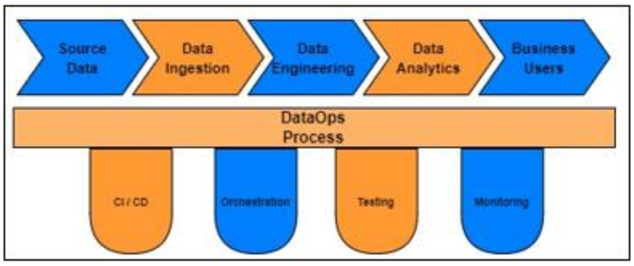

Figure 2: DataOps Process

4. Real Time Data Processing

Real - time data processing is essential to modern business functions. The process helps detect and manage fraud. Real - time data processing uses various tools and technologies to facilitate data ingestion and processing. The tools used for data ingestion are:

1) **Apache NiFi: -** Apache NiFi is an open real - time data ingestion platform designed to handle data transfer between different source and target systems. This allows you to automate the movement and processing of data. The platform provides a visual interface that allows you to design data pipelines and consume them from different sources.
2) **Apache Kafka: -** It is a distributed platform that enables high throughput and fault tolerant messaging, it facilitates the real time data retrieval process. The process also allows you to transform the flow and use of data from multiple sources to enable a comprehensive approach.
3) **Apache Flume: -** It is a platform that offers an ingestion mechanism for gathering, organizing, and transmitting extensive streaming data, such as events and log files from various sources into a centralized data warehouse. Flume is a highly reliable, distributed and configurable tool. It is mainly designed to copy streaming data (log data) from various servers on the Internet to the Hadoop Distributed File System (HDFS).
4) **Amazon Kinesis: -** It is a product of Amazon Web Services and is applicable for data processing and real - time streaming. It can help with data ingestion, analysis and processing at any available scale [12].

4.1. Data Preparation and Enrichment process

Real - time data processing for fraud detection requires established steps to assist with handling and managing data preparation and enrichment. These steps and practices allow the distribution and automation of procedures to ensure high quality data for analysis. Some key practices during this process include the following aspects:

1) **Data Cleaning: -** This helps to identify and remove errors, inconsistencies, duplicates and missing entries from the data to increase consistency and the quality of the data.
2) **Data Transformation: -** This process converts data into a format suitable for analysis. The process consists of data collection and normalization, as well as categorization [13].
3) **Data Enrichment: -** This process combines internal data from internal sources with disparate data from other internal systems or third - party data from external sources. Enriched data is a valuable asset to any organization as it becomes more useful and relevant. These additional sources may include geospatial information, demographic data, and historical transaction data.
4) **Metadata Management: -** This is an approach that provides a complete understanding of the data, taking into account the source, quality and lineage of the data. This allows information traceability.

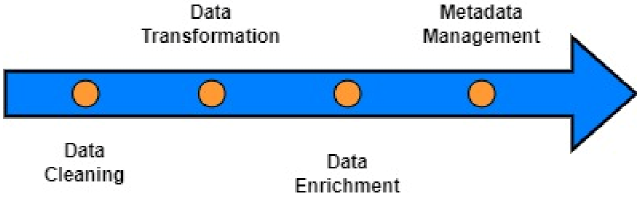

Figure 3: Data Preparation and Enrichment Steps

4.2. Streaming Analytics for Fraud Detection

Stream analysis allows continuous analysis of data streams to detect anomalies. The techniques that apply in streaming analytics for fraud detection include:

1) **Machine Learning Models: -** Unsupervised and supervised machine learning mechanisms can identify patterns of fraudulent activity through training. The use of real - time data sources ensures the detection and classification of transaction anomalies. This also allows the prioritization of changes based on the possibility of fraud [14].
2) **Rule - based Systems: -** This mechanism defines the main conditions for the alert. The approach responds when specific patterns and behaviors are observed, including detecting fraud patterns through predefined historical patterns, thresholds and heuristics.
3) **Complex Event Processing (CEP): -** CEP works with data flow analysis to detect sequences that indicate fraudulent behavior. These CEP approaches help detect fraud schemes involving multiple transactions or parties.
4) **Stream Processing Frameworks: -** Frameworks like Apache Flink provide resources for real - time analysis on various data streams. They support real - time data filtering, clustering and pattern recognition.

Businesses have an opportunity to ensure they can use streaming analytics to improve early detection and response to fraudulent activity. This system allows data ingestion, preparation and enrichment with key tools that ensure accurate identification of fraudulent activities. Adopting these practices ensures that companies can perform well and achieve the most outstanding results by generating and ensuring consistent results in each indicator.

4.3. Elasticsearch for Real Time Analysis

Elasticsearch is a great addition to real - time analytics in fraud detection systems. The platform can be used for indexing and research activities. To enable indexing, the platform primarily uses a very efficient structure called an index, which is field - based and makes data searchable and retrievable. This approach provides an instrumental capability to identify key data sources for the fraud detection system. However, the platform allows searching with key questions, which allows searching through large volumes of unstructured data and using different formats to obtain the tremendous advantage of data management [16]. The platform is also scalable, which allows the data to be used on several nodes and to manage a large volume of data, entering several requests at the same time. The scalability element ensures that fraud detection adapts to the growing nature of data volumes and requests, while maintaining the same data performance model. This implies that it can be used to carry out many research models, such as geospatial research, respond to the increasing growth of information and achieve the reliable goal of the management of information development. However, the development of real - time indexing makes the system easy to use, where information can be obtained in a fraction of a minute, ensuring a stable management of the entire information system. Elasticsearch plays a key role in improving real - time fraud detection with its extensive capabilities to leverage exceptional data management and engagement approaches. Elasticsearch is of great importance for fraud detection systems. First, it allows you to detect fraudulent activity in real time using patterns and key information. Real - time data indexing and analysis allows companies to easily recognize and mitigate fraudulent activity, resulting in fewer losses and reputational damage [17]. The platform also has the ability to recognize patterns and detect anomalies. This is done through historical information on fraudulent activity, which allows detection and prevention of fraud earlier before it escalates to a full - scale attack. The system is also scalable and can be applied to different storage models. The overall approach is based on the potential and ability to enable consistent results in fraud detection and management of organizational needs.

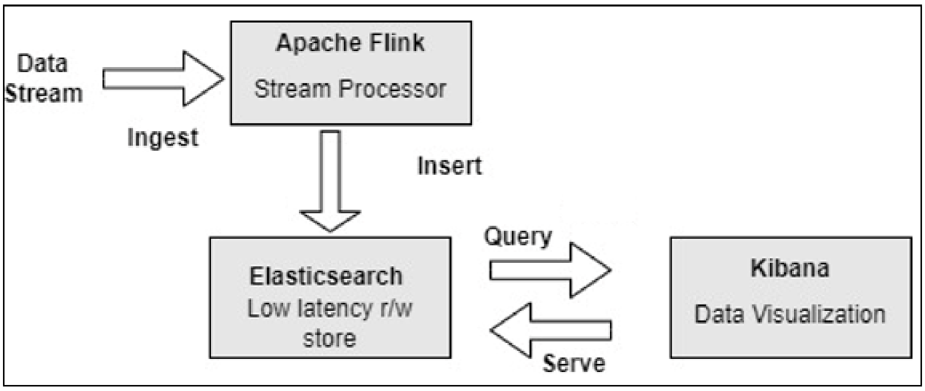

Figure 4: Elasticsearch for Real Time Data Handling

4.4. Alerting and Response

Fraud detection systems should work with alert and response systems, which ensure that an organization is aware of potential fraudulent activity and can take action as quickly as possible. Real - time notifications are an essential step to advance solutions and achieve the best result to transform and determine the best resources to achieve realistic results. Different alert and response approaches can be used to allow the system to have greater traction in addressing relevant issues. Primarily, the use of threshold - based alerts ensures that the key conditions under which an alert is triggered are met. Having thresholds in the form of system anomalies, number of transactions and user behaviors creates a system approach for details and administers the value management of the appropriate items. Therefore, the activation of these alerts is essential to define a new level of fraud management as soon as it occurs. In addition, to enable real - time alerts, the attention of observers in Elasticsearch ensures that the functionality can manage alerts with specifications for key changes. Watchers allow organizations to define their actions to group and trigger changes by indicating the occurrence of specific events and patterns. The alert can also be made through integration with an external notification mechanism that provides faster information. Using tools such as SMS and e - mail, an appropriate level of screening of basic identification and registration requirements can be carried out. Rapid response and insights through integrated platforms encourage the creation of mechanisms that support and address progressive fraud detection, which is an essential addition to modern businesses. The implementation of the rapid response helps to minimize financial losses and encourage the management of the company to achieve stable results in the detection of unauthorized transactions and those whose limits exceed normal financial limits. It also indicates the ability of an entity to have the flexibility to implement compliance requirements and regulatory provisions that ensure successful management and treatment of its requirements as necessary [18]. In this case, the lines of compliance and fraud detection are handled appropriately in the company, ensuring a successful consideration of the underlying variables to obtain and guarantee consistent results in all aspects. Therefore, these regulations are taken into account and the appropriate systems are implemented to obtain the appropriate result in all complaints. A quick response to fraud detection marks the company's opportunity to take steps to improve its reputational integrity. The key infrastructure to assist in the detection, processing and communication of any fraudulent activity is an indicator of the fundamental characteristics that influence the nature of business claims management. Using key risk management strategies to identify fraud, inform on its opportunities and mitigate its impact on the business restores confidence in the business, helping to create an outstanding address of the relevant enabling factors in all aspects. Therefore, the rapid response to the detection of fraud allows the critical adaptation of the needs of the organization to face the risks and maintain its reputation to resist such situations.

5. Compliance Security

Compliance and security of fraud detection systems are essential to ensure the management of sensitive data. The use of strict regulations and rules in the management of resilient data platforms is a notable step that allows the correct management of fraud detection in companies to the extent necessary. To enable compliance with regulatory requirements, access controls must be implemented, ensuring that appropriate and appropriate steps are taken to allow the resilient data platform to have authorized data management for each category of employees. Role - based access control (RBAC) and multi - factor authentication help follow appropriate guidelines to manage and process individual steps to achieve desired meaningful results. In particular, the use of access controls must be combined with complete audit trails and log management for which information is processed, with a key distinction so that any influential information can be traced in the platform. Essentially, the management of these advances ensures that data audits can highlight any changes in the system and encourage instrumental modeling to achieve the right proportions of the required value. Greater security for fraud detection can be achieved by enabling encryption at rest and in transit. Processing information in these two distinct categories helps to obtain the ability to influence to face the challenges in the system. Encryption sets the right tone and precedent to assert better information management mechanisms, emphasizing that only authorized individuals have the ability to use and process information at will. Thus, these advances help to critically advance the value of growth and management of information as expected. Therefore, the use of appropriate security measures to enable encryption promotes the progressive processing of information as desired. Fraud detection can be handled by various encryption techniques, which help improve system compliance and security requirements. The use of anonymization creates an extraordinary step in which privacy is respected, ensuring that identifiable personal information is processed in the appropriate framework and the ability to achieve the desired reliable result. This approach reduces the risk of a data breach or even the use of sensitive information in analytical activities. In addition, the use of tokenization creates unique identifiers for data, helping to build and create best - in - class sensitive data management during data processing and analysis on elastic data platforms. The use of homomorphic encryption is also essential to process calculations that do not require decryption of information. This approach helps to manage data privacy and security during data processing and analysis techniques. Homomorphic cryptography allows incredible measures of information processing and obtaining reliable results that ensure that real - time data can value the representation of information at any cost.

6. Conclusion

Fraud detection is an important factor for organizations today. Fraud detection can be managed using real - time data through DataOps on Elastic platforms, RiskOps and fraud detection software to provide exceptional traction for managing unauthorized changes and activities in a network. DataOps enables data management to ensure that reliable information is provided as part of regular practices, providing guidance on anomalies and aspects of data that may trigger alerts. The use of Elastic platforms provides a critical attraction for the scalability and flexibility that allow the management of information on a large set of data. The ability to manage these data sets creates sustainable options for addressing management paths based on critical information to provide real - time alerts, analyze information, and work toward correct fraud detection in each system. Thus, the use of DataOps on Elastic platforms allows the implementation of real - time data management and analysis, which allows a critical model of real - time alerts to emerge for information.

References

[1] C. Onwubiko Fraud matrix: a morphological and analysis - based classification and taxonomy of fraud Comput. Secur. (2020)

[2] Matheus Severino et al. Machine learning algorithms for fraud prediction in property insurance: empirical evidence using real - world microdata Mach. Learn. Appl. (2021)

[3] Y. Sahin et al. A cost - sensitive decision tree approach for fraud detectionExpert Syst. Appl. (2013)

[4] S. Zoldi Using anti - fraud technology to improve the customer experienceComput. Fraud Secur. (2015)

[5] B. Baesens et al. robROSE: a robust approach for dealing with imbalanced data in fraud detection Stat. Methods Appl. (2021)

[6] D. Choi and K. Lee, "An artificial intelligence approach to financial fraud detection under IoT environment: A survey and implementation, " Security and Communication Networks, 2018.

[7] S. V. S. S. Lakshmi and S. D. Kavilla, "Machine learning for credit card fraud detection system, " International Journal of Applied Engineering Research, vol.13, no.24, pp.16819 - 16824, 2018.

[8] A. Saputra, "Fraud detection using machine learning in e - commerce, " International Journal of Advanced Computer Science and Applications, vol.10, no.9, 2019.

[9] Aparna Krishna Bhat, “Application And Impact of Artificial Intelligence in Financial Decision Making”, Int J Sci Res Sci Eng Technol, vol.11, no.5, pp.57–63, Sep.2024, doi: 10.32628/IJSRSET2411417.

[10] K. K. Voruganti, "Leveraging DataOps Principles for Efficient Data Management in Cloud Environments, " Journal of Technological Innovations, vol.4, no.4, 2023.

[11] "ElasticSearch: Everything you need to know about this software", https: //datascientest. com/en/elasticsearch - everything - you - need - to - know

[12] "Streaming Data Solutions on AWS", https: //d0. awsstatic. com/whitepapers/whitepaper - streaming - data - solutions - on - aws - with - amazon - kinesis. pdf

[13] Aparna Krishna Bhat, "YugabyteDB Data Migration Best Practices Guide", https: //medium. com/[at]bhat_aparna1/yugabytedb - data - migration - best - practices - guide - 174d2ecbd35e

[14] J. Sreemathy et al., "Overview of ETL tools and talend - data integration, " in 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Mar.2021, vol.1, pp.1650 - 1654.

[15] M. Bendechache et al., "Modelling and simulation of ElasticSearch using CloudSim, " in 2019 IEEE/ACM 23rd International Symposium on Distributed

Simulation and Real Time Applications (DS - RT), Oct.2019, pp.1 - 8.

[16] A. Deniz, M. M. Elömer, and A. A. Aydin, "A comparison of Apache Solr and Elasticsearch technologies in support of large - scale data analysis, " Gümüşhane Üniversitesi Fen Bilimleri Dergisi, vol.13, no.2, pp.386 - 404, 2023

[17] Z. Lu et al., "On the auto - tuning of elastic - search based on machine learning, " in Proceedings of the 2020 1st International Conference on Control, Robotics and Intelligent System, Oct.2020, pp.150 - 156

[18] M. Siering, J. Muntermann, and M. Grčar, "Design principles for robust fraud detection: The case of stock market manipulations, " 2021.

[19] G. Calderon, G. Del Campo, E. Saavedra, and A. Santamaria, "Management and monitoring IoT networks through an elastic stack - based platform, " in 2021 8th International Conference on Future Internet of Things and Cloud (FiCloud), Aug.2021, pp.184 - 191.

[20] H. Atwal and H. Atwal, "Dataops technology, " in Practical DataOps: Delivering Agile Data Science at Scale, 2020, pp.215 - 247.