World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(REVIEW ARTICLE)

Check for updates

Leveraging ETL automation and data analytics to drive a reduction in software development bottlenecks: Lessons from Global South

David Akokodaripon [1,*] and Godwin Nzeako [2]

[1] *Take-Blip, Belo-Horizonte, Brazil.*
[2] *Independent Researcher, Finland.*

World Journal of Advanced Research and Reviews, 2023, 19(03), 1650-1662

Publication history: Received on 26 July 2023; revised on 14 September 2023; accepted on 17 September 2023

Article DOI: https://doi.org/10.30574/wjarr.2023.19.3.1796

Abstract

This study explores how ETL (Extract, Transform, Load) automation and data analytics can alleviate software development bottlenecks, with a focus on the global South. The aim is to uncover strategies that enhance software quality, shorten delivery times, and cut costs by leveraging data-driven methodologies. The research employs a comprehensive review of existing literature, analyzing contemporary tools, technologies, and best practices. It emphasizes the role of ETL automation in integrating data analytics into the software development lifecycle, showcasing how these processes drive efficiency and sustainability.

The study's key findings highlight that ETL automation plays a critical role in improving data consistency, accuracy, and processing speed, thereby enhancing the reliability of software solutions. Data analytics, on the other hand, enables predictive modeling and real-time decision-making, crucial for proactively managing development challenges. The research further stresses the value of incorporating circular economy principles and indigenous practices to promote sustainability and cultural relevance in software development. These findings advocate for a holistic approach in the global South, where embracing regulatory compliance, climate resilience, and stakeholder engagement can significantly benefit software development practices.

Concluding that the adoption of ETL automation and data analytics is essential for resolving software development bottlenecks, the study recommends that organizations integrate advanced technologies while aligning their strategies with both local and global standards. This approach not only streamlines the development process but also fosters the creation of innovative, sustainable software solutions. The study suggests future research directions, including the integration of AI and machine learning into ETL processes, the establishment of comprehensive policy frameworks, and the promotion of interdisciplinary collaboration to further advance the field.

Keywords: ETL automation; Data analytics; Software development bottlenecks; Global South; Sustainability; Predictive modeling

1. Introduction

The increasing complexity and dynamism of software development pose significant challenges, particularly when it comes to mitigating development bottlenecks. These challenges are more pronounced in the global South, where technological infrastructure, access to advanced tools, and skilled expertise may vary widely. In this context, ETL (Extract, Transform, Load) automation and data analytics have emerged as powerful approaches to streamline software development processes. By harnessing ETL automation, organizations can enhance data processing efficiency, thus reducing delays and optimizing the development pipeline (Buinwi et al., 2022). This integration of ETL processes is

* Corresponding author: David Akokodaripon

especially critical in handling large volumes of data that must be processed quickly and accurately, reducing the manual effort and errors often associated with complex software projects.

In the Global South, software development encounters distinct challenges, such as restricted access to advanced technology, financial limitations, and a shortage of specialized technical expertise (Kabanda, Tanner & Kent, 2018). These challenges can impede the efficiency of software development projects, leading to delays, higher costs, and diminished quality. Gade (2021) highlights the importance of operational improvements, such as cost management strategies and data-driven decision-making, in addressing these issues. ETL automation, for example, provides a systematic method for handling large datasets, reducing reliance on manual processes. This not only accelerates data processing but also lowers the risk of human error, resulting in more precise and dependable software products. When integrated with data analytics, ETL automation delivers actionable insights, empowering organizations to make informed decisions, optimize workflows, and improve overall software quality.

Data analytics plays a critical role in driving competitive advantage in software development by enabling predictive models and data-driven decision-making. Buinwi et al. (2022) highlight the importance of data analytics in identifying bottlenecks and optimizing resource allocation and process flows. Through predictive analytics, organizations can anticipate potential delays and take proactive measures to mitigate them. This is particularly crucial in the global South, where resource limitations require strategic prioritization and allocation to ensure optimal use. For example, data analytics can identify underperforming areas in the development pipeline, allowing teams to reallocate resources efficiently and focus on high-impact tasks. By employing data analytics, organizations can achieve significant gains in efficiency and productivity, thereby reducing the time and cost associated with software development.

In addition to enhancing operational efficiency, data analytics contributes to improved software quality. By analyzing historical data and performance metrics, development teams can identify patterns and trends that may affect the software's performance and reliability. This allows for continuous improvement and refinement of the software development process. Moreover, data analytics can be used to assess user behavior and feedback, providing valuable insights into user preferences and requirements. This user-centric approach ensures that the developed software aligns with market demands and expectations, further driving the product's success.

The implementation of ETL automation and data analytics in software development is heavily shaped by the legal and regulatory environment. Muhammad et al. (2022) highlight that the growing reliance on digital transactions and software systems demands robust legal frameworks to ensure data security and compliance. In the Global South, where digital infrastructure is still developing, establishing these frameworks remains a significant challenge. Aligning the adoption of ETL automation and data analytics with evolving legal requirements is crucial for sustainable and secure software development. Adherence to data protection regulations, such as the General Data Protection Regulation (GDPR) and other regional laws, is vital for safeguarding user data and maintaining trust. Organizations must carefully navigate these legal complexities while integrating ETL and analytics solutions to ensure their software development practices remain efficient and compliant.

The circular economy, which focuses on sustainability and resource optimization, is increasingly being incorporated into software development practices. Kabanda, Tanner and Kent (2018) suggest that embracing circular economy principles can enhance resource efficiency, reduce waste, and improve sustainability in software development. ETL automation plays a pivotal role in this approach by streamlining the extraction, transformation, and loading of data, enabling organizations to manage data in ways that minimize waste and encourage the reuse and recycling of data resources. For instance, automating data workflows reduces the energy and resources needed for data processing, contributing to a more sustainable software development lifecycle. Furthermore, ETL processes facilitate the integration of legacy systems with modern technologies, minimizing the need for resource-intensive system overhauls and supporting the sustainable evolution of software solutions.

The evolving trade and industrial policies in the global South have a significant impact on software development practices. Buinwi and Buinwi (2022) note that government policies and international trade agreements increasingly recognize the role of digital technologies and data analytics in economic growth. By aligning software development practices with these policies, organizations can leverage ETL automation and data analytics to reduce bottlenecks and enhance productivity. This alignment, however, requires a strategic approach that takes into account the unique economic, technological, and regulatory contexts of the region. For instance, trade policies that encourage technology transfer and investment in digital infrastructure can facilitate the adoption of ETL automation and data analytics in software development. Similarly, policies that support skill development and capacity building can help bridge the technical expertise gap, enabling organizations in the global South to effectively implement and benefit from ETL and analytics solutions.

Furthermore, the integration of ETL automation and data analytics can support the development of software solutions tailored to the specific needs of the global South. For example, ETL processes can be used to integrate data from diverse sources, such as local market data, user feedback, and regulatory requirements, into a unified framework. This enables the development of software that is not only technologically advanced but also culturally and contextually relevant. Data analytics can further enhance this by providing insights into user behavior, market trends, and regulatory changes, allowing for the continuous adaptation and improvement of software products.

This review paper aims to explore how ETL automation and data analytics can be leveraged to address software development bottlenecks, drawing lessons from the global South. The objective is to provide a comprehensive understanding of the strategies and outcomes associated with integrating these technologies into the software development lifecycle. By examining the intersection of ETL automation, data analytics and software development within the specific context of the global South, this paper seeks to identify best practices, challenges, and future research directions. The scope includes analyzing the impact of regulatory frameworks, technological infrastructure and circular economy principles on software development efficiency and sustainability. Through this exploration, the paper aims to contribute to the broader discourse on how ETL automation and data analytics can drive innovation and efficiency in software development, particularly in regions facing unique technological and infrastructural challenges.

2. Conceptual Framework of ETL Automation and Data Analytics

The integration of Extract, Transform, Load (ETL) automation and data analytics is essential for effective data management, particularly in complex digital environments. ETL automation simplifies the process of extracting, transforming, and loading data from multiple sources into a centralized warehouse, allowing organizations to manage large volumes of data more efficiently (George, 2023). This automation is vital for ensuring data quality, consistency, and availability—key factors for successful data analytics and informed decision-making. In today's digital age, where data generation and processing are continuous, ETL automation enables organizations to handle extensive datasets with ease, reducing manual effort and minimizing errors.

The exponential growth of digital data has posed significant challenges in its management and security (Miryala & Gupta, 2022). ETL automation helps mitigate these challenges by standardizing data formats and centralizing storage, which is particularly vital in sectors such as healthcare and banking, where accurate and timely data analysis is crucial. For example, in banking, ETL processes facilitate the secure extraction, transformation, and loading of transaction data, ensuring analytics can be performed without compromising data integrity. Furthermore, data analytics enhances this workflow by delivering insights that enable real-time decision-making and ensure adherence to privacy regulations (Munagandla, Dandyala & Vadde, 2022).

The integration of ETL automation and data analytics is crucial for ensuring compliance with global data privacy regulations. Laws such as the General Data Protection Regulation (GDPR) mandate stringent data management practices, including anonymization and secure storage of personal information. ETL automation facilitates compliance by automating the transformation and secure storage of data (Miryala & Gupta, 2022). Meanwhile, data analytics tools support ongoing monitoring of data usage, enabling organizations to proactively detect and address potential privacy risks, thus maintaining compliance with regulatory requirements.

ETL automation frameworks play a crucial role in promoting digital inclusion, especially in regions with limited connectivity. By streamlining the processing and analysis of data on digital access, organizations can identify underserved communities and craft targeted strategies to enhance connectivity (Miryala & Gupta, 2022). Moreover, data analytics offers valuable insights into user behavior, enabling the development of digital services tailored to the diverse needs of various populations. This method not only advances digital inclusion but also ensures that privacy and security remain central to decision-making processes.

In cybersecurity, ETL automation plays a vital role in ensuring secure data management. With the expansion of digital banking services, the demand for secure data processing has become increasingly critical (Bibi, 2022). ETL processes facilitate the secure extraction, transformation, and loading of data into protected environments, mitigating the risk of breaches. Complementing this, data analytics enables real-time monitoring of transactional data, allowing for the prompt identification and response to security threats.

Emerging technologies such as artificial intelligence (AI) and big data greatly enhance ETL and data analytics frameworks. AI-powered ETL tools automate intricate data transformations, boosting efficiency and accuracy in data processing (Nookala et al., 2020). Meanwhile, big data analytics enables organizations to process and analyze massive datasets in real time, delivering actionable insights for strategic decision-making. For instance, in public health

campaigns, AI and big data can be leveraged to analyze trends and support targeted interventions. However, these advancements must be carefully managed to address privacy concerns, particularly when handling sensitive data, ensuring ethical and compliant data practices.

2.1. Software Development in the Global South: An Overview

Software development in the Global South presents a unique mix of challenges and opportunities, shaped by the region's diverse economic, cultural, and technological contexts. Creating software that is both technically robust and contextually appropriate is essential, particularly given the demand for sustainable business models and the often limited digital infrastructure (Di Vaio et al., 2020). Cybersecurity and data privacy are especially critical as reliance on digital platforms grows. Industries such as finance and healthcare, which handle highly sensitive data, require software with advanced security features to protect user information and comply with relevant regulations (Rao & Deebak, 2023). This focus on security aligns with circular economy principles, where secure data management is integral to optimizing resource use and minimizing waste (Di Vaio et al., 2020).

The integration of ETL (Extract, Transform, Load) automation and data analytics into software development is essential for the Global South, where resource constraints and infrastructure challenges are prevalent. ETL automation streamlines the efficient extraction and processing of data from diverse sources, supporting decision-making in sectors where data collection often presents difficulties (Muhammad et al., 2022). In Nigeria's banking sector, data analytics and ETL processes play a crucial role in managing operational risks by enabling timely data analysis to support better decision-making (Nwafor, Nwafor & Onalo, 2019). This efficient approach to data management is vital in settings that demand both operational excellence and adherence to regulatory standards.

Cybersecurity considerations significantly shape software development in the Global South. In sustainable supply chain management, software must tackle cybersecurity challenges to ensure data integrity and defend against cyberattacks (Rao & Deebak, 2023). Similarly, implementing smart grid technologies for sustainable energy requires software with a strong focus on security to prevent potential disruptions (Alotaibi et al., 2020). Consequently, software developers in the Global South must incorporate robust security measures to safeguard critical infrastructure and maintain data integrity.

Data privacy and security introduce additional challenges, particularly in fields like environmental research, where safeguarding sensitive data is essential (Rao & Deebak, 2023). Developers must contend with diverse data privacy laws across countries, ensuring compliance to avoid legal complications and protect their reputation. Blockchain technology has proven to be an effective solution for securely managing digital transactions, providing a transparent approach to preserving data integrity (Muhammad et al., 2022).

In addition to data processing and cybersecurity, ETL automation in software development across the Global South enables large-scale data analytics, fostering innovation and informing policy-making. For instance, in public health, software systems with automated ETL processes support real-time monitoring and targeted interventions, which are essential in regions with constrained healthcare resources (Miryala & Gupta, 2022). This data-driven approach enhances the ability to manage public health challenges more effectively.

The adoption of circular economy principles is transforming data and resource management in software development across the Global South (Di Vaio et al., 2020). Sustainable software solutions are increasingly designed to optimize resource utilization, reduce waste, and extend product lifecycles. Achieving this requires adaptable software capable of integrating with diverse data sources, making ETL automation and data analytics indispensable for enabling sustainable business practices.

Despite significant progress, challenges such as limited technology access, a shortage of skilled professionals, and varying levels of digital literacy remain. However, initiatives centered on capacity building, knowledge sharing, and promoting digital inclusion are actively addressing these barriers (Miryala & Gupta, 2022). By encouraging technological innovation and investing in skill development, the Global South can strengthen its software development capabilities and drive sustainable economic growth.

2.2. Role of ETL Automation in Reducing Software Development Bottlenecks

ETL (Extract, Transform, Load) automation plays a vital role in addressing software development bottlenecks by simplifying data integration, processing, and transformation. Managing tasks like data extraction, transformation, and loading into data warehouses often creates delays and complexities in software development. By automating these repetitive processes, ETL automation enables development teams to focus on more critical stages of the development

cycle (Muhammad et al., 2022). This is especially significant in the Global South, where software environments frequently encounter resource and infrastructure constraints.

ETL automation greatly improves data processing efficiency in software development. In data-intensive sectors like healthcare, it facilitates the rapid extraction and transformation of vast amounts of data into actionable insights (Rao & Deebak, 2023). By minimizing the risk of errors associated with manual data handling, ETL automation enhances the quality of software systems. For example, it ensures the accurate and secure processing of sensitive data, such as patient records, safeguarding data integrity and privacy. As a result, ETL automation not only accelerates development but also supports the creation of reliable and secure software solutions.

ETL automation is pivotal for ensuring regulatory compliance in software development. As data privacy and security gain prominence under regulations such as GDPR, organizations must implement strong data management practices (Muhammad et al., 2022). ETL automation facilitates compliance by processing and storing data in accordance with privacy standards, reducing the risk of breaches and unauthorized access. This is especially critical in the Global South, where diverse legal frameworks make regulatory adherence more complex. By offering a standardized approach, ETL automation ensures consistent data privacy and security across various jurisdictions.

From a cybersecurity standpoint, ETL automation is pivotal in protecting software systems. As cyber threats continue to evolve, software must incorporate strong mechanisms to safeguard data (Alotaibi et al., 2020). ETL automation strengthens security by facilitating the secure extraction, transformation, and loading of data. For instance, in waste management technologies, ETL automation ensures the secure processing of sensitive information throughout its lifecycle (Alotaibi et al., 2020). By automating these security protocols, ETL minimizes the risk of data breaches, enhancing the resilience of software systems against cyber threats.

In educational settings, ETL automation is equally important for developing and deploying eLearning systems. For instance, in Nigerian tertiary institutions, the adoption of eLearning technology has presented challenges related to data management and scalability (Olatubosun et al., 2015). ETL automation addresses these challenges by seamlessly integrating diverse data sources such as student information systems, learning management systems, and digital libraries. This integration ensures data is consistently updated and accurately processed, enhancing the user experience and supporting the effective delivery of educational content. Additionally, automating data processes helps maintain data integrity and security, crucial for protecting student information and complying with educational data privacy regulations.

Moreover, ETL automation supports data-driven decision-making in various sectors. In public health, for instance, software systems analyzing spatial patterns of diseases like mother-to-child HIV transmission rely on processing large datasets efficiently. ETL automation streamlines this process by automating the extraction and transformation of data from multiple sources, generating timely and accurate insights essential for public health interventions.

ETL automation effectively resolves bottlenecks associated with data quality and consistency. Variations in data formats and structures often lead to integration challenges and system errors. By transforming data into a standardized format before loading it into target systems, ETL automation eliminates these issues (Muhammad et al., 2022). This ensures that software applications operate with consistent, accurate, and reliable data, reducing errors and system malfunctions. As a result, ETL automation improves software quality and speeds up development by minimizing the need for manual data cleaning and validation.

2.3. Data Analytics as a Tool for Optimizing Software Development

Data analytics plays a crucial role in optimizing software development by providing deep insights into development processes, identifying bottlenecks, and enabling data-driven decision-making. Utilizing techniques such as statistical modeling, predictive analytics, and machine learning, data analytics enhances both the efficiency and quality of software development (Makinde & Fasoranbaku, 2011). This strategy enables teams to proactively address challenges, optimize workflows, and make informed decisions to enhance both the development process and the quality of the final software product (Boppana, 2019).

A key application of data analytics in software development is performance monitoring. By analyzing performance data, developers can identify patterns that may indicate inefficiencies or bottlenecks (Makinde & Fasoranbaku, 2011). For example, maximum depth classifiers can help analyze software system depth distributions, pinpointing critical areas where performance might be compromised. This proactive approach allows developers to address issues efficiently, thereby enhancing software stability and performance.

Predictive analytics is also vital, as it uses historical data to forecast potential software failures, project delays, or resource allocation needs (Makinde & Fasoranbaku, 2018). Autoregressive Integrated Moving Average (ARIMA) models, explored by Makinde and Fasoranbaku (2018) effectively predict software performance trends. By anticipating these trends, teams can better allocate resources, plan maintenance activities, and mitigate risks, ensuring optimal software performance.

Machine learning algorithms significantly improve software development by automating anomaly detection and error correction. Belciug and Gorunescu (2014) highlight the use of backpropagation procedures in medical diagnosis, which can be adapted for identifying software errors. By analyzing development data, machine learning facilitates automated debugging, reducing the risk of unnoticed errors and enhancing overall software quality.

Data analytics also significantly improves project management by offering insights into task completion times, resource usage, and team performance. Omolofe et al. (2022) suggest using multivariate control charts to monitor project performance over time, allowing managers to detect deviations and take corrective actions promptly. Data analytics also strengthens software security by examining usage patterns to identify vulnerabilities or cyber threats (Rao & Deebak, 2023). This real-time analysis facilitates early detection of threats and enables swift responses, thereby maintaining the integrity of software systems.

In the Global South, where software development faces limited resources and infrastructural constraints, data analytics is crucial for optimizing processes and maximizing resources (Makinde & Fasoranbaku, 2013). By leveraging data-driven strategies, development teams can create high-quality software solutions that address regional needs, even in environments with limited resources (Balasundaram, Sathiyaseelan & Zirkler, 2023).

2.4. Challenges of Implementing ETL Automation and Data Analytics in the Global South

Implementing ETL (Extract, Transform, Load) automation and data analytics in the Global South presents several challenges rooted in technical, infrastructural, and socio-economic factors. One of the primary difficulties is the lack of optimal data collection and processing strategies, which can significantly affect the quality and reliability of analytics (Olamide, Fasoranbaku & Adebola 2021). Optimal experimental designs, such as A-optimal and E-optimal designs discussed by Olamide, Fasoranbaku and Adebola (2021), are essential for efficient data collection and analysis. However, the application of these advanced statistical models often requires expertise and resources that may be scarce in the Global South. This limitation hampers the ability to implement robust ETL automation processes that rely on high-quality data for accurate analytics.

Data quality is a crucial factor that impacts the effectiveness of ETL automation. In the Global South, incomplete or inconsistent data can pose a significant challenge, necessitating advanced techniques to ensure data integrity during the extraction and transformation phases (Aladeniyi, Olowofeso & Fasoranbaku, 2009). Techniques for handling incomplete data, such as those proposed by Aladeniyi, Olowofeso and Fasoranbaku (2009), are vital in this context. However, the application of such methods can be complex and resource-intensive, often requiring specialized knowledge in statistical modeling and data processing. This complexity can result in higher implementation costs and longer development times, thus affecting the adoption of ETL automation and data analytics in software development projects.

Another challenge is the need for optimal design and estimation methods to support data analytics in software development. Fasoranbaku and Daramola (2016) emphasize the importance of optimal experimental designs for estimating parameters in linear regression models, which are often used in software development analytics. In the Global South, the implementation of such optimal designs is complicated by resource constraints, including limited access to computational tools and software necessary for performing complex statistical analyses. Moreover, the scarcity of trained professionals skilled in these advanced methodologies can further exacerbate the challenge, making it difficult to leverage ETL automation effectively.

Infrastructure limitations in the Global South, including inadequate computational resources and data storage capabilities, pose additional obstacles to implementing ETL automation. Effective ETL processes require reliable and scalable infrastructure to handle large volumes of data efficiently. However, in many regions of the Global South, there is a shortage of such infrastructure, which can lead to inefficiencies in data processing and analytics workflows (Adebola, Fasoranbaku & Kupolusi, 2020). The optimal design of data analytics models, such as those for Poisson regression and quadratic logistic models Adebola, Fasoranbaku and Kupolusi (2020) demands significant computational power. Without the necessary infrastructure, it becomes challenging to implement these models at scale, hindering the full potential of ETL automation in optimizing software development.

Socio-economic factors significantly influence the challenges of implementing ETL automation and data analytics. The high initial costs of acquiring advanced ETL tools, investing in training, and upgrading infrastructure can be a major obstacle for organizations in the Global South, especially small and medium-sized enterprises (SMEs) (Otemoyolo, 2021). These financial limitations hinder the adoption of comprehensive ETL automation solutions capable of meeting the diverse demands of software development projects in the region. Furthermore, resistance to adopting new technologies, often stemming from a lack of awareness or understanding of the benefits of ETL automation and data analytics, further slows progress.

In the context of data analytics, estimating non-linear regression parameters using denoised variables, as suggested by Fasoranbaku and Alabi (2016), is essential for enhancing the accuracy of predictive models. However, the complexity involved in such estimation techniques presents a barrier to their widespread adoption in the Global South, where there may be limited expertise in advanced statistical methods. This limitation can lead to suboptimal implementation of data analytics, resulting in less effective software development processes. Therefore, addressing the challenges of implementing ETL automation and data analytics in the Global South requires a multi-faceted approach that considers technical, infrastructural, and socio-economic factors.

2.5. Impact on Software Quality, Delivery Time, and Cost

The implementation of ETL (Extract, Transform, Load) automation and data analytics has a transformative effect on software quality, delivery timelines, and cost efficiency, particularly in the Global South. By automating data integration processes, software quality is significantly enhanced as human errors in data handling and transformation are minimized (Thumburu, 2021). Since the accuracy of software outputs is closely tied to the quality of input data, ETL automation systematically cleanses, transforms, and loads data, reducing the risk of errors in the final product. This streamlined data processing approach improves the reliability and overall performance of software systems.

Effective management of multicollinearity and other data complexities is essential in software development to ensure high-quality outcomes. Gudivada, Apon, and Ding (2017) introduced a Bayesian ridge estimator for logistic regression to address multicollinearity, offering more accurate parameter estimates. Integrating this approach into ETL processes enhances data transformation stages by enabling the handling of complex datasets more efficiently. By incorporating advanced statistical methods like Bayesian estimation, ETL pipelines improve the precision and robustness of software algorithms. Ultimately, this integration strengthens ETL automation, contributing to the development of resilient and high-quality software products capable of managing data variability.

ETL automation also has a significant impact on delivery time. Traditional data integration methods often involve manual data extraction, transformation, and loading, which are time-consuming and prone to errors. By automating these processes, development teams can accelerate software development cycles, enabling faster delivery of software products (Adewole & Fasoranbaku, 2021). The use of Bayesian estimation techniques for determining optimal hyperparameters, as discussed by Adewole and Fasoranbaku (2021), can further streamline the ETL process by facilitating automated tuning of model parameters. This automation reduces the time developers spend on manual adjustments, allowing them to focus on other critical aspects of software development, thus speeding up the overall project timeline.

Cost reduction is a significant benefit of implementing ETL automation and data analytics. Automating data integration processes reduces the reliance on manual intervention, thereby lowering labor expenses. Moreover, automated ETL processes improve data accuracy and consistency, resulting in fewer errors and less rework during software development (Kothandapani, 2021). This decrease in errors translates to reduced maintenance costs, as high-quality software requires less extensive debugging and patching after deployment. Additionally, advanced ETL tools optimize resource utilization, cutting computational costs associated with processing large volumes of data.

In the Global South, where resources are often constrained, the cost-effectiveness of ETL automation offers significant advantages. Salamkar and Allam (2019) stress the importance of strategies that promote agile software development and efficient resource utilization. By integrating ETL automation, software development teams in these regions can boost productivity despite limited resources. This increased efficiency not only lowers development costs but also enhances the accessibility and affordability of software solutions for a wider range of users and organizations.

The influence of ETL automation and data analytics extends beyond the initial development phase to the ongoing maintenance and improvement of software systems. Automated data processing pipelines allow software to adapt seamlessly to changes in data structures and sources, ensuring scalability and flexibility over time. This adaptability is especially vital in dynamic environments where data requirements shift rapidly (Endres, Endres & Berg, 2018). For

instance, in sustainable architectural solutions for affordable housing, ETL automation supports the continuous monitoring and analysis of data on energy consumption, building materials, and occupancy trends. This integration enables the development of more efficient and sustainable software solutions for urban planning and management (Endres, Endres & Berg, 2018).

2.6. Tools and Technologies for ETL Automation and Data Analytics

The integration of ETL (Extract, Transform, Load) automation and data analytics in software development relies on various tools and technologies that streamline data processing and enable real-time insights. A key feature of these tools is their ability to handle large-scale data efficiently, ensuring software systems are built on reliable datasets (Olamide, Fasoranbaku & Adebola, 2021). D-optimal designs for polynomial Poisson regression models provide a robust statistical framework to optimize data analysis, managing complex data relationships that ETL processes encounter in software development.

Advanced ETL tools such as Apache NiFi, Talend, and Microsoft SQL Server Integration Services (SSIS) streamline the extraction, transformation, and loading of data from diverse sources (Patel & Patel, 2020). These platforms enable the management of complex workflows, ensuring data consistency and integrity throughout the software development lifecycle. By reducing the need for manual intervention, these tools improve software quality and accelerate delivery timelines by minimizing errors in data management.

Data analytics platforms like Apache Spark, Hadoop, and Power BI play a critical role in processing and analyzing large datasets to derive actionable insights (Ikegwu et al., 2022). These tools support advanced analytics capabilities, including predictive modeling, trend analysis, and performance monitoring. For instance, integrating Apache Spark with ETL pipelines enhances data processing speed and enables real-time analytics. Such integration is vital for agile decision-making in software development, helping to identify bottlenecks and optimize resource allocation, thereby improving project efficiency.

The implementation of smart grids and renewable energy systems highlights the importance of leveraging advanced data analytics to manage complex operations (Bhattarai et al., 2019). Similarly, in software development, integrating ETL automation with machine learning enables the prediction of system behavior by analyzing historical and real-time data. This predictive capability facilitates proactive adjustments, improving software performance and reliability.

Cloud-based ETL and analytics platforms such as AWS Glue, Google BigQuery, and Azure Data Factory provide scalable solutions for managing data in the cloud (Anoshin et al., 2020). These services are especially beneficial for organizations in the Global South, where infrastructure limitations can create challenges. By automating data workflows across distributed environments, cloud-based ETL tools enable efficient data processing and analysis regardless of location. This scalability accommodates the expanding data requirements of software projects, allowing development teams to concentrate on innovation.

Integrating renewable energy technologies with ETL automation underscores the potential for advancing sustainable software development (Endres, Endres & Berg, 2018). Energy-efficient ETL processes and data analytics tools help minimize operational costs and reduce energy consumption, aligning software development with sustainability objectives. By optimizing server utilization and streamlining data processing tasks, ETL automation fosters a more environmentally responsible approach to software development.

2.7. Lessons Learned and Best Practices

The integration of ETL automation and data analytics into software development has yielded valuable insights and best practices, particularly in optimizing workflows, improving decision-making, and gaining a competitive edge (Housley et al., 2014). A key takeaway is the critical role of predictive models and data-driven strategies in streamlining processes and mitigating bottlenecks. ETL tools equipped with analytics capabilities allow organizations to efficiently process large datasets, derive actionable insights, and make informed decisions that enhance software quality and performance. This data-focused approach meets the evolving demands of modern software development, emphasizing agility and adaptability as essential for success.

Integrating circular economy principles into ETL processes has proven instrumental in promoting sustainable and resource-efficient software development (Kabanda, Tanner & Kent, 2018). By combining ETL automation with data analytics, organizations can maximize resource efficiency, reduce waste, and improve the sustainability of their software solutions. This approach encourages a culture of continuous improvement, where data analytics is utilized to monitor and optimize development processes, ensuring resources are managed effectively throughout the software

lifecycle. Adopting circular economy strategies in ETL automation not only supports environmental sustainability but also unlocks opportunities for cost savings and increased operational efficiency.

A key insight is the importance of thoroughly understanding trade and industrial policies when integrating ETL automation and data analytics into software development (Housley et al., 2014). Regions in the Global South, for instance, encounter distinct challenges such as regulatory compliance, technological infrastructure limitations, and dynamic market conditions. A deep understanding of local policy frameworks is crucial for addressing these challenges and aligning ETL processes with regional needs. Best practices involve collaboration with policymakers, industry stakeholders, and regulatory authorities to ensure that ETL automation and data analytics adhere to regulations and contribute to overarching economic and industrial goals.

The integration of indigenous practices into modern software development frameworks has been recognized as a best practice in the context of the Global South (Irani et al., 2010). Indigenous knowledge systems, which prioritize sustainability, community engagement, and resource efficiency, provide valuable perspectives for improving software development processes. By embedding indigenous principles into ETL automation and data analytics, organizations can create software solutions that are culturally relevant, socially inclusive, and environmentally sustainable. This approach fosters the development of software systems that are both technically advanced and attuned to the socio-cultural and economic needs of their target users.

Technological advancements, especially in data analytics, play a pivotal role in creating a competitive edge in software development (Housley et al., 2014). Employing advanced analytics tools, such as machine learning algorithms and predictive models, allows organizations to uncover patterns, foresee challenges, and enhance software performance. Key practices in this domain include ongoing monitoring and analysis of software metrics to inform decision-making and optimize processes. By utilizing data analytics, organizations can preemptively resolve bottlenecks, improve resource allocation, and elevate the quality and reliability of their software products.

Promoting a culture of innovation and continuous learning is essential for the effective implementation of ETL automation and data analytics (Endres, Endres & Berg, 2018). This requires fostering an environment that supports experimentation, knowledge sharing, and the adoption of emerging technologies and practices. Organizations should focus on training and development programs to enhance their teams' expertise in data analytics and ETL automation. By fostering a culture of data-driven decision-making and continuous improvement, organizations can stay ahead of industry trends and deliver high-quality software solutions that meet the evolving needs of their users (Akindote et al., 2023).

2.8. Future Trends and Research Directions

The integration of ETL automation and data analytics in software development is advancing, with future trends focusing on more sophisticated and adaptable technologies. A key trend is the incorporation of artificial intelligence (AI) and machine learning (ML) into ETL processes, enabling advanced data extraction, transformation, and loading capabilities (Housley et al., 2014). AI-powered ETL tools will leverage data patterns to automate complex transformations, streamline workflows, and detect bottlenecks with greater efficiency. These advancements will be especially valuable in the Global South, where resource limitations necessitate innovative approaches to enhance efficiency and productivity.

An emerging trend in software development is the emphasis on regulatory compliance and the integration of trade policy considerations. Housley et al. (2014) highlight the challenges and opportunities posed by international trade policies, underscoring the need for a more sophisticated approach to compliance. Future research is expected to investigate how ETL automation can be tailored to navigate complex regulatory landscapes, particularly in regions such as Cameroon, where trade policies heavily influence software markets. Efforts in this area will aim to create frameworks that embed policy considerations into ETL processes, ensuring software systems remain aligned with evolving regulations and trade agreements.

Sustainability and climate resilience are increasingly pivotal in software development, as noted by Singh and Goyal (2023). Future advancements in ETL automation are expected to incorporate sustainable practices, such as enhancing energy efficiency in data centers and minimizing the carbon footprint of software development operations. Furthermore, research will likely focus on leveraging data analytics to support climate-resilient software systems, particularly in regions vulnerable to climate change. For instance, predictive analytics can model the impacts of environmental factors on software infrastructure, facilitating the creation of more robust and adaptable systems.

The influence of indigenous knowledge and practices on the future of software development is profound and cannot be underestimated. Cole (2020) highlights the significance of incorporating indigenous architectural principles to enhance climate resilience in urban housing. Similarly, software development stands to gain from research that integrates local knowledge systems into ETL and data analytics frameworks. Future investigations will focus on how indigenous practices can inform the design and implementation of software systems, particularly in areas like user interface design, data privacy, and community engagement. This approach aspires to produce software solutions that are both technologically sophisticated and culturally attuned, effectively addressing the needs of local communities.

Collaboration and interdisciplinary research are essential for advancing future trends in software development and data management. As these fields grow increasingly complex, the need for joint efforts among experts from disciplines such as computer science, public policy, environmental science, and sociology will intensify (Housley et al., 2014). Interdisciplinary research enables the creation of comprehensive solutions that address the technical, regulatory, and societal dimensions of ETL automation and data analytics. By promoting cross-disciplinary collaboration, future research can support the development of software systems that are more resilient, adaptable, and aligned with global sustainability objectives.

3. Conclusion

This study set out to explore how ETL automation and data analytics can address software development bottlenecks, with a particular focus on the unique challenges faced in the global South. The aim was achieved by examining the role of ETL automation in streamlining data processing and enhancing software quality, delivery time, and cost-efficiency. Through an analysis of various tools, technologies, and practices, this study demonstrated how ETL processes contribute to more efficient data management and decision-making, ultimately reducing development delays and improving software products.

Key findings highlight the significance of incorporating ETL automation and data analytics into software development workflows. ETL automation improves data consistency and accuracy, which directly enhances software quality and reduces delivery times. Moreover, data analytics allows for a data-driven approach to software development, enabling predictive modeling and real-time decision-making. This has been shown to significantly reduce costs, particularly in resource-constrained environments like the global South. The integration of circular economy principles and indigenous practices was identified as crucial for achieving sustainability and contextual relevance in software solutions.

The study concludes that implementing ETL automation and data analytics in software development is essential for overcoming bottlenecks and achieving high-quality, sustainable software solutions. To this end, it is recommended that organizations adopt a holistic approach that includes regulatory compliance, cultural adaptation, and climate resilience considerations. Additionally, fostering collaboration and stakeholder engagement is crucial for creating software that aligns with both local needs and global standards.

Overall, this study provides a comprehensive framework for understanding and leveraging ETL automation and data analytics in software development. It underscores the need for continued research into advanced technologies, sustainable practices, and policy integration to further enhance the efficiency and impact of software development processes. By implementing the recommendations and insights provided, organizations can navigate the complexities of software development, particularly in the global South, and deliver innovative, effective, and sustainable software solutions.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Adebola, F.B., Fasoranbaku, O.A. and Kupolusi, J.A., (2020). On prediction error variance to determining optimal design for two variable quadratic logistic model. Cogent Mathematics & Statistics, 7(1), p.1853888. https://doi.org/10.1080/25742558.2020.1853888

[2] Adewole, A.I. and Fasoranbaku, O.A., (2021). Determination of Quantile Range of Optimal Hyperparameters Using Bayesian Estimation. Tanzania Journal of Science, 47(3), pp.988-998. DOI:10.4314/tjs.v47i3.10

[3] Akindote, O.J., Egieya, Z.E., Ewuga, S.K., Omotosho, A. and Adegbite, A.O., (2023). A review of data-driven business optimization strategies in the US economy. International Journal of Management & Entrepreneurship Research, 5(12), pp.1124-1138. https://doi.org/10.51594/ijmer.v5i12.681

[4] Aladeniyi, O.B., Olowofeso, O.E. and Fasoranbaku, O.A., (2009). On new techniques for testing incomplete data using regression models. The Pacific Journal of Science and Technology, 10, pp.149-159.

[5] Alotaibi, I., Abido, M.A., Khalid, M. and Savkin, A.V., (2020). A comprehensive review of recent advances in smart grids: A sustainable future with renewable energy resources. Energies, 13(23), p.6269. https://doi.org/10.3390/en13236269

[6] Anoshin, D., Foshin, D., Storchak, R. and Ireton, X., (2020). Azure Data Factory Cookbook: Build and manage ETL and ELT pipelines with Microsoft Azure's serverless data integration service. Packt Publishing Ltd.

[7] Balasundaram, S., Sathiyaseelan, A. and Zirkler, M., (2023). Jugaad in organizational settings: exploring the Jugaad leadership competencies. Asian Business & Management, 22(5), pp.1877-1912. https://doi.org/10.1007/s13132-024-01862-8

[8] Belciug, S. and Gorunescu, F., (2014). Error-correction learning for artificial neural networks using the Bayesian paradigm. Application to automated medical diagnosis. Journal of Biomedical Informatics, 52, pp.329-337. https://doi.org/10.1016/j.jbi.2014.07.013

[9] Bhattarai, B.P., Paudyal, S., Luo, Y., Mohanpurkar, M., Cheung, K., Tonkoski, R., Hovsapian, R., Myers, K.S., Zhang, R., Zhao, P. and Manic, M., (2019). Big data analytics in smart grids: state-of-the-art, challenges, opportunities, and future directions. IET Smart Grid, 2(2), pp.141-154. https://doi.org/10.1049/iet-stg.2018.0261

[10] Bibi, P., (2022). Artificial Intelligence in Cybersecurity: Revolutionizing Database Management for Enhanced Protection.

[11] Boppana, V.R., (2019). Implementing Agile Methodologies in Healthcare IT Projects. SSRN, 4987242.

[12] Buinwi, A., Okatta, C. G., Johnson, E., Buinwi, J. A., and Tuboalabo, A. (2022). Enhancing trade policy education: A review of pedagogical approaches in public administration programs. International Journal of Applied Research in Social Sciences, 6(6).

[13] Buinwi, E., Buinwi, J. A., Okatta, C. G., and Johnson. (2022). Leveraging business analytics for competitive advantage: Predictive models and data-driven decision making. International Journal of Management & Entrepreneurship Research, 6(6), 997-2014.

[14] Buinwi, J. A., and Buinwi, U. (2022). Challenges and Opportunities in International Trade Policy Implementation: Insights from the Cameroonian Ministry of Trade. International Journal of Management & Entrepreneurship Research, 6(7).

[15] Buinwi, J. A., and Buinwi, U. (2022). The evolution of trade and industrial policies: Lessons from Cameroon. International Journal of Advanced Economics, 6(7), 319-339.

[16] Cole, R.J., (2020). Navigating climate change: Rethinking the role of buildings. Sustainability, 12(22), p.9527. https://doi.org/10.3390/su12229527

[17] Di Vaio, A., Palladino, R., Hassan, R. and Escobar, O., (2020). Artificial intelligence and business models in the sustainable development goals perspective: A systematic literature review. Journal of Business Research, 121, pp.283-314.

[18] Endres, H., Endres and Berg, (2018). Adaptability Through dynamic capabilities. Wiesbaden: Springer Fachmedien Wiesbaden.

[19] FASORANBAKU, O. and DARAMOLA, G., (2018). Towards a better estimation of the parameters of linear regression models: The optimal designed experiment approach. International Journal of New Technology and Research, 4(5), p.263062.

[20] Fasoranbaku, O.A. and Alabi, O., (2016). Estimating Non-linear Regression Parameters Using Denoised Variables.

[21] Fasoranbaku, O.A. and Makinde, O.S., (2011). Behavioural pattern of causality parameter of autoregressive moving average model. Journal of the Nigerian Association of Mathematical Physics, 19.

[22] Gade, K.R., (2021). Data-Driven Decision Making in a Complex World. Journal of Computational Innovation, 1(1).

[23] George, J., (2023). Harnessing the power of real-time analytics and reverse ETL: Strategies for unlocking data-driven insights and enhancing decision-making. Available at SSRN 4963391. http://dx.doi.org/10.2139/ssrn.4963391

[24] Gudivada, V., Apon, A. and Ding, J., (2017). Data quality considerations for big data and machine learning: Going beyond data cleaning and transformations. International Journal on Advances in Software, 10(1), pp.1-20.

[25] Housley, W., Procter, R., Edwards, A., Burnap, P., Williams, M., Sloan, L., Rana, O., Morgan, J., Voss, A. and Greenhill, A., (2014). Big and broad social data and the sociological imagination: A collaborative response. Big Data & Society, 1(2), p.2053951714545135.

[26] Ikegwu, A.C., Nweke, H.F., Anikwe, C.V., Alo, U.R. and Okonkwo, O.R., (2022). Big data analytics for data-driven industry: a review of data sources, tools, challenges, solutions, and research directions. Cluster Computing, 25(5), pp.3343-3387. https://doi.org/10.1007/s10586-022-03568-5

[27] Irani, L., Vertesi, J., Dourish, P., Philip, K. and Grinter, R.E., (2010), April. Postcolonial computing: a lens on design and development. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 1311-1320). https://doi.org/10.1145/1753326.1753522

[28] Kabanda, S., Tanner, M. and Kent, C., (2018). Exploring SME cybersecurity practices in developing countries. Journal of Organizational Computing and Electronic Commerce, 28(3), pp.269-282. https://doi.org/10.1080/10919392.2018.1484598

[29] Kothandapani, H.P., (2021). Integrating Robotic Process Automation and Machine Learning in Data Lakes for Automated Model Deployment, Retraining, and Data-Driven Decision Making.

[30] Makinde, O.S. and Fasoranbaku, O.A., (2011). Identification of optimal autoregressive integrated moving average model on temperature data. Journal of Modern Applied Statistical Methods, 10, pp.718-729. https://doi.org/10.56801/10.56801/v10.i.581

[31] Makinde, O.S. and Fasoranbaku, O.A., (2018). On maximum depth classifiers: depth distribution approach. Journal of Applied Statistics, 45(6), pp.1106-1117. https://doi.org/10.1080/02664763.2017.1342783

[32] Makinde, O.S., Adegbie, K.S. and Fasoranbaku, O.A., (2013). Mathematical modelling of two-step exothermic reactions with and without reactants consumption. FUTA J. Res. Sci, 1, pp.44-53.

[33] Miryala, N.K. and Gupta, D., (2022). Data security challenges and industry trends. IJARCCE International Journal of Advanced Research in Computer and Communication Engineering, 11(11), pp.300-309.

[34] Muhammad, T., Munir, M.T., Munir, M.Z. and Zafar, M.W., (2022). Integrative cybersecurity: merging zero trust, layered defense, and global standards for a resilient digital future. International Journal of Computer Science and Technology, 6(4), pp.99-135.

[35] Munagandla, V.B., Dandyala, S.S.V. and Vadde, B.C., (2022). The future of data analytics: trends, challenges, and opportunities. Revista de Inteligencia Artificial en Medicina, 13(1), pp.421-442.

[36] Nookala, G., Gade, K.R., Dulam, N. and Thumburu, S.K.R., (2020). Automating ETL Processes in Modern Cloud Data Warehouses Using AI. MZ Computing Journal, 1(2).

[37] Nwafor, C.N., Nwafor, O.Z. and Onalo, C., (2019). The use of business intelligence and predictive analytics in detecting and managing occupational fraud in Nigerian banks. Journal of Operational Risk, Forthcoming.

[38] Olamide, E.I., Adebola, F.B. and Fasoranbaku, O.A., (2021). D-OPTIMAL DESIGNS FOR POLYNOMIAL POISSON REGRESSION MODELS ON CONSTRAINED DESIGN SPACE. Annals. Computer Science Series, 19(1).

[39] Olamide, E.I., Adebola, F.B. and Fasoranbaku, O.A., (2021). Weighted Statistical A-optimal Designs for Two-variable Poisson Regression Model with Application to Fertility Studies. Tanzania Journal of Science, 47(3), pp.1303-1309. DOI:10.4314/tjs.v47i3.37

[40] Olamide, E.I., Fasoranbaku, O.A. and Adebola, F.B., (2021). E-optimal Experimental Designs for Poisson Regression Models in Two and Three Variables. Tanzania Journal of Science, 47(3), pp.999-1006. DOI:10.4314/tjs.v47i3.11

[41] Olatubosun, O., Olusoga, F. and Abayomi, F., (2015). Diabetes Diagnosis with Maximum Covariance Weighted Resilience Back Propagation Procedure. British Journal of Mathematics & Computer Science, 6(5), pp.381-393. https://doi.org/10.9734/BJMCS/2015/14871

[42] Olatubosun, O., Olusoga, F. and Samuel, O., (2015). Adoption of eLearning technology in Nigerian tertiary institution of learning. British Journal of Applied Science and Technology, 10(2), pp.1-15. https://doi.org/10.9734/BJAST/2015/18434

[43] Olopha, P.O., Fasoranbaku, A.O. and Gayawan, E., (2021). Spatial pattern and determinants of sufficient knowledge of mother to child transmission of HIV and its prevention among Nigerian women. Plos one, 16(6), p.e0253705.

[44] Omolofe, O.T., Adegoke, N.A., Adeoti, O.A., Fasoranbaku, O.A. and Abbasi, S.A., (2022). Multivariate control charts for monitoring process mean vector of individual observations under regularized covariance estimation. Quality Technology & Quantitative Management, 19(3), pp.277-298. https://doi.org/10.1080/16843703.2021.1948952

[45] Otemoyolo, M.A., (2021). Challenges and Critical Success of Factors of Business Intelligence for Small and Medium Enterprises in the Republic of Congo. Colorado Technical University.

[46] Patel, M. and Patel, D.B., (2020). Progressive growth of ETL tools: A literature review of past to equip future. Rising Threats in Expert Applications and Solutions: Proceedings of FICR-TEAS 2020, pp.389-398. https://doi.org/10.1007/978-981-15-6014-9_45

[47] Rao, P.M. and Deebak, B.D., (2023). Security and privacy issues in smart cities/industries: technologies, applications, and challenges. Journal of Ambient Intelligence and Humanized Computing, 14(8), pp.10517-10553. https://doi.org/10.1007/s12652-022-03707-1

[48] Salamkar, M.A. and Allam, K., (2019). Data Lakes Vs. Data Warehouses: Comparative Analysis on When to Use Each, With Case Studies Illustrating Successful Implementations. Distributed Learning and Broad Applications in Scientific Research, 5.

[49] Singh, S. and Goyal, M.K., (2023). Enhancing climate resilience in businesses: The role of artificial intelligence. Journal of Cleaner Production, 418, p.138228. https://doi.org/10.1016/j.jclepro.2023.138228

[50] Thumburu, S.K.R., (2021). Optimizing Data Transformation in EDI Workflows. Innovative Computer Sciences Journal, 7(1).