International Journal of Science and Engineering, 2024; 1(4):47-62
ISSN: *2998-4874*
Doi: 10.62304/ijse.v1i04.188

FUTURE TRENDS IN SQL DATABASES AND BIG DATA ANALYTICS: IMPACT OF MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE

Siful Islam[1]

[1]Graduate Researcher, Master of Science in Management Information Systems, College of Business, Lamar University, Texas, US

Email: sislam13@lamar.edu

https://orcid.org/0009-0009-0592-2059

Keywords

SQL Databases
Big Data Analytics
Machine Learning
Artificial Intelligence
Query Optimization
Predictive Analytics
Data Security

***Received**: 11, May, 2024*
***Accepted**: 02, August, 2024*
***Published**: 06, August,2024*

ABSTRACT

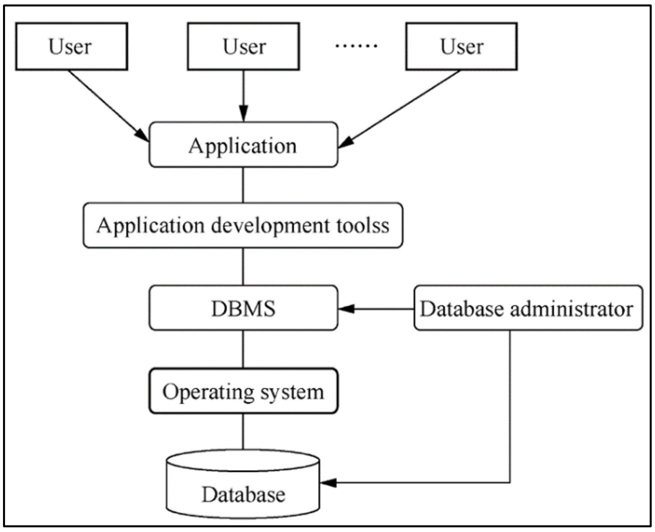
This study systematically reviews the integration of machine learning (ML) and artificial intelligence (AI) into SQL databases and big data analytics, highlighting significant advancements and emerging trends. Using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, a comprehensive review of 60 selected articles published between 2010 and 2023 was conducted. The findings reveal substantial improvements in query optimization through ML algorithms, which adapt dynamically to changing data patterns, reducing processing times and enhancing performance. Additionally, embedding ML models within SQL databases facilitates real-time predictive analytics, streamlining workflows, and improving the accuracy and speed of predictions. AI-driven security systems provide proactive and real-time threat detection, significantly enhancing data protection. The development of hybrid systems that combine relational and non-relational databases offers versatile and efficient data management solutions, addressing the limitations of traditional systems. This study confirms the evolving role of AI and ML in transforming data management practices and aligns with and extends previous research findings.

1 Introduction

The exponential growth of data in recent years has necessitated advancements in database management systems, particularly SQL databases, to handle increasingly complex and voluminous datasets (Hou et al., 2016; Hsu et al., 2014; Indu et al., 2017). SQL databases, traditionally known for their structured query capabilities, have evolved to meet the demands of modern big data applications. This evolution is marked by the integration of big data analytics, enabling businesses to derive meaningful insights from vast amounts of data (Grolinger et al., 2013; Liao et al., 2016). As organizations generate and collect more data from diverse sources, the scalability and performance of SQL databases have become critical factors in their adoption and implementation (Miranskyy et al., 2017). In addition, the intersection of SQL databases and big data analytics is further enhanced by the advent of machine learning (ML) and artificial intelligence (AI). These technologies bring unprecedented capabilities to data processing and analysis, allowing for more sophisticated query optimization, predictive analytics, and anomaly detection (Hosseinzadeh et al., 2018). AI and ML algorithms can analyze patterns within the data, providing predictive insights that were previously unattainable with traditional database systems (Badampudi et al., 2015). This integration facilitates a more dynamic and responsive approach to data management, where real-time insights can significantly influence business strategies and operations (Alzoubi et al., 2020; Badampudi et al., 2015).

One of the key benefits of incorporating ML and AI into SQL databases is the enhancement of query optimization. Traditional query optimization relies on predefined rules and historical performance data, which can be limiting in rapidly changing data environments (Alzoubi & Ahmed, 2019; Roy, 2022). In contrast, AI-driven optimization techniques continuously learn from the database workload, adapting to changes and improving query performance over time (Badampudi et al., 2015). This adaptability is crucial for maintaining the efficiency of database operations as the volume and variety of data continue to grow (Liao et al., 2016). Another significant impact of AI and ML on SQL databases is in the realm of data security. AI-driven security measures can detect unusual patterns and potential threats more effectively than traditional methods (Roy, 2022). Machine learning models can be trained to recognize anomalies in data access and usage, providing early warnings of security breaches and helping to mitigate risks (Grolinger et al., 2013). This proactive approach to security is essential in an era where cyber threats are becoming increasingly sophisticated and pervasive (Alshurideh, 2022).

Figure 1: Database System



The future of SQL databases is also shaped by the trend towards hybrid systems that combine the strengths of relational and non-relational databases. These systems are designed to leverage the structured querying capabilities of SQL databases with the flexibility and scalability of big data frameworks like Hadoop and Spark (Elkington, 1997). By integrating these technologies, organizations can benefit from the robust data management features of SQL while also harnessing the power of big data analytics to handle unstructured and semi-structured data (Joy, Abdulla, et al., 2024). This convergence is expected to drive further innovations in data processing and analysis, making it possible to address the complex data challenges of the future (Joy, Rahman, et al., 2024).

Doi: 10.62304/ijse.v1i04.188

Correspondence: Siful Islam
Graduate Researcher, Master of Science in Management Information Systems, College of Business, Lamar University, Texas, US

Email: sislam13@lamar.edu

OPEN ACCESS

Doi: 10.62304/ijse.v1i04.188

The objective of this paper is to investigate the future trends in SQL databases and big data analytics, with a particular focus on the impact of machine learning (ML) and artificial intelligence (AI) on these technologies. Specifically, the study aims to explore how ML and AI are enhancing query optimization, predictive analytics, and data security within SQL databases. Additionally, the paper seeks to examine the role of AI-driven automation in database management and its potential to reduce operational costs and improve efficiency. Furthermore, the research aims to analyze the convergence of SQL databases with big data frameworks, highlighting the development of hybrid systems that combine the strengths of relational and non-relational databases. Through this comprehensive analysis, the paper intends to provide insights into the technological advancements and industry practices that will shape the future of SQL databases and big data analytics, driven by the capabilities of ML and AI.

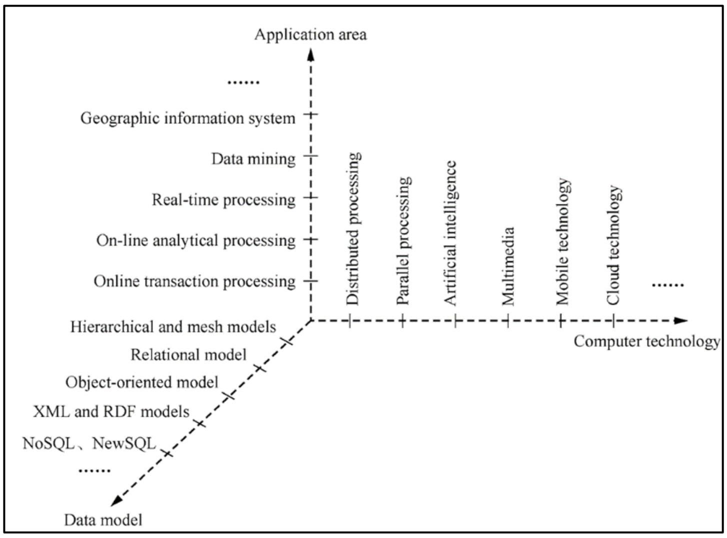
2 Literature Review:

The integration of SQL databases with big data analytics has advanced significantly, addressing scalability and performance issues inherent in traditional systems. Studies by Rauf et al. (2024) emphasize the necessity of combining SQL with big data frameworks to manage large-scale data effectively. The infusion of machine learning (ML) and artificial intelligence (AI) into SQL databases has revolutionized query optimization and data retrieval, as highlighted by Md Mahfuzur et al. (2024). AI-driven techniques not only enhance the efficiency of database operations but also improve data security by identifying and mitigating potential threats (Joy, Rahman, et al., 2024). These developments underscore the evolving role of SQL databases in supporting complex analytical tasks and real-time decision-making in modern data environments.

2.1 SQL Databases and Big Data Analytics

SQL databases have been foundational to data management due to their robust structure and efficient query processing capabilities. Lee, Azmi, et al. (2022) and AlHamad et al. (2022) highlight that SQL databases, designed to handle structured data through a relational model, are ideal for applications requiring consistency and reliability. Traditional SQL databases, such as PostgreSQL and MySQL, utilize structured query language to manage and manipulate data, offering transactional support and ensuring data integrity (Roy, 2022). However, the emergence of big data—characterized by high volume, velocity, and variety—has challenged the effectiveness of traditional SQL databases (Petcu et al., 2013). The rise of big data analytics has transformed data processing and decision-making, enabling the extraction of valuable insights from vast and complex datasets that traditional SQL databases struggle to handle on their own (Stanescu et al., 2016). This shift has led to the integration of SQL databases with big data technologies like Hadoop and Spark, leveraging the strengths of both relational and non-relational models (Yoon et al., 2016). Ali et al. (2022) discuss how this integration allows for scalable data processing and real-time analytics, crucial for modern business applications. Lee and Kang (2015) and McCoy (2017) further emphasize that this hybrid approach enables organizations to retain the transactional benefits of SQL while benefiting from the scalability and flexibility of big data frameworks.

Figure 2: Applications and related technologies and models of Database Systems



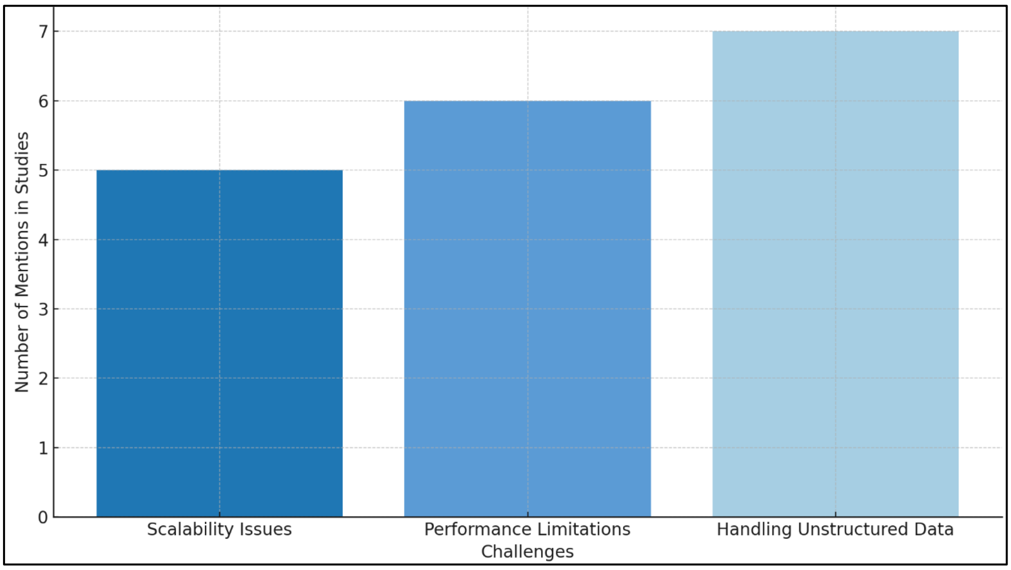
The integration of SQL databases with big data analytics frameworks has led to significant advancements in data management and analysis. By combining the structured data capabilities of SQL with the unstructured data processing power of big data technologies, organizations can achieve a more comprehensive data strategy (Radwan & Farouk, 2021). Petersen et al. (2015) illustrate how modern architectures facilitate seamless querying of both structured and unstructured data, enhancing overall data analytics efficiency and effectiveness. This integration is further enhanced by applying machine learning (ML) and artificial intelligence (AI) technologies, which bring advanced analytical capabilities to SQL databases. Alomari and Noaman (2019) and Rocha et al. (2015) demonstrate that AI-driven analytics enable real-time insights and predictive modeling, essential for data-driven decision-making processes. The convergence of SQL databases with big data analytics also addresses several traditional data management system limitations. Baralis et al. (2017) and Mondol (2021) discuss how traditional SQL databases face challenges in scalability and handling diverse data types. Integrating big data frameworks allows these systems to overcome such limitations by distributing data processing tasks across multiple nodes, thus enhancing performance and scalability (Ivanov & Pergolesi, 2019). Additionally, Al Ali (2021) highlight the role of AI and ML in automating complex data management tasks, reducing operational overhead, and improving data security. The synthesis of these technologies marks a significant step forward in the evolution of data management, enabling organizations to harness the full potential of their data assets.

2.2 Challenges in Traditional SQL Databases

Traditional SQL databases, while powerful and reliable, face several notable challenges that limit their efficacy in handling modern data demands. Scalability is a primary concern, as these databases are designed to manage structured data on a single server or a limited number of nodes (Ghazal et al., 2021). As data volumes grow exponentially, the ability of SQL databases to scale efficiently becomes strained, leading to performance bottlenecks (Alomari & Noaman, 2019). Research by Alomari et al. (2014)highlights the inherent difficulties in scaling traditional relational databases horizontally, compared to the relative ease with which NoSQL and big data frameworks can expand. Additionally, Mondol (2021) note that while vertical scaling (adding more power to existing machines) is possible, it is often cost-prohibitive and ultimately limited by the hardware capabilities. These scalability issues can significantly hinder the performance of SQL databases in large-scale data environments, where rapid growth and high data throughput are the norms (Ivanov & Pergolesi, 2019).

Performance limitations are closely tied to the scalability issues in traditional SQL databases. As the volume and complexity of data increase, the efficiency of query processing and transaction management can degrade (Rathika, 2019). Mondol (2021) discusses how the fixed schema design and the need for complex joins and indexing in SQL databases can lead to substantial

Figure 3: Challenges in Traditional SQL Databases



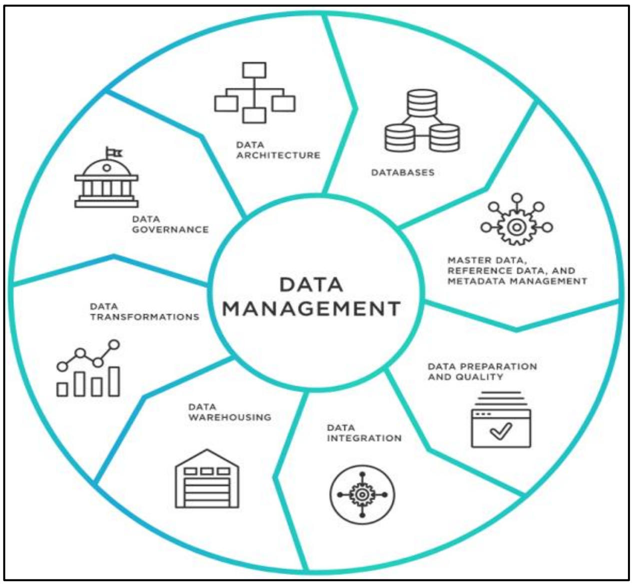
performance overhead, especially when dealing with large datasets. Al Ali (2021) argue that the rigid structure of SQL databases, while beneficial for data integrity and consistency, can impede performance when handling the dynamic and varied nature of big data. Furthermore, traditional SQL databases often struggle with the real-time processing requirements demanded by modern applications, as their architectures are not optimized for low-latency data access and high-throughput transactions (Hanaysha et al., 2021). Studies by Inglot and Kozioł (2016) and Khan et al. (2022) also point out that the performance of SQL databases can be adversely affected by the need to maintain ACID (Atomicity, Consistency, Isolation, Durability) properties, which adds additional processing overhead and complexity (Shamim, 2022).

Handling unstructured data presents another significant challenge for traditional SQL databases. These databases are designed to work with structured data, organized into tables with predefined schemas (Lee, Romzi, et al., 2022). However, the rise of big data has brought an influx of unstructured data types, such as text, images, and social media content, which do not fit neatly into relational tables (Lee & Ahmed, 2021). Ivanov and Pergolesi (2019) emphasize that SQL databases lack the flexibility to efficiently store and query unstructured data, necessitating the use of additional tools and frameworks to manage these data types. This limitation is highlighted by Khan et al. (2022), who show that integrating unstructured data into a traditional SQL environment often requires significant preprocessing and transformation, which can be both time-consuming and resource-intensive. Additionally, Lee et al. (2021) note that while extensions and adaptations like JSON support have been introduced to SQL databases, they still fall short in providing the seamless handling and performance optimization offered by native NoSQL solutions. As a result, organizations often turn to hybrid systems that combine SQL and NoSQL technologies to better manage diverse data types and optimize performance (Ordonez, 2010; Pokorny, 2013).

2.3 Integration of SQL with Big Data Frameworks

The integration of SQL databases with big data frameworks, such as Hadoop and Apache Spark, has become essential for managing the complexities of modern data environments. Architectures combining SQL and big data leverage the strengths of both relational and non-relational models, creating a more robust data management system (Sharma et al., 2018). Hadoop's distributed storage and processing capabilities efficiently handle large-scale unstructured data, while SQL databases provide reliable transactional support and structured querying (Ayub & Ali, 2018). Apache

Figure 4: Integration of SQL with Big Data Frameworks



Spark enhances this integration by offering in-memory processing and advanced analytics, enabling faster data processing and real-time insights (Ribas et al., 2015). According to Ghazal (2021), these integrated architectures facilitate seamless data movement and querying across diverse sources, providing a unified platform for comprehensive data analysis. This combination allows organizations to harness the benefits of SQL's structured data management and the scalability of big data technologies, ensuring efficient handling of vast datasets.

The advantages of integrating SQL databases with big data frameworks are significant and multifaceted. One major benefit is the ability to perform large-scale data analytics without compromising the transactional integrity provided by SQL databases (AlHamad et al., 2021). This integration enables complex analytical queries on vast datasets stored in Hadoop's distributed file system (HDFS), while maintaining SQL's consistency and reliability in transactions (Ahlers & Wilde, 2017).

Furthermore, integrated systems support both batch processing and real-time analytics, enhancing versatility for various business needs (Alshurideh et al., 2022). Apache Spark's in-memory processing capabilities significantly reduce data operation latency, boosting SQL query performance on big data (Ali et al., 2021). Anand and Rao (2016) note that this synergy allows organizations to gain real-time insights and make data-driven decisions more efficiently. Additionally, frameworks like Apache Hive and Cloudera Impala enable SQL queries directly on HDFS data, preserving SQL's familiarity while extending its utility to large-scale analytics (Rautmare & Bhalerao, 2016). Pokorny (2013) emphasize that these integrations simplify data workflows by reducing the need for extensive data transformations and migrations. The incorporation of machine learning and graph processing libraries in Apache Spark further broadens the analytical capabilities of SQL databases (Mehmood & Anees, 2019). This synthesis of SQL with big data frameworks marks a significant advancement in data processing, addressing traditional system limitations and unlocking new possibilities for managing complex data environments (Ayub & Ali, 2018; Li & Manoharan, 2013; Sharma et al., 2018).

2.4 Role of Machine Learning in SQL Databases

Machine learning (ML) has significantly enhanced the capabilities of SQL databases, particularly in the realm of query optimization. Traditional query optimization techniques in SQL databases rely heavily on predefined rules and historical performance data, which can be insufficient in handling complex and dynamic data environments (Ribas et al., 2015). ML algorithms, on the other hand, analyze patterns within the database workload and adaptively optimize queries over time, leading to substantial performance improvements (Ghazal, 2021). These algorithms learn from data and query execution statistics to predict the most efficient execution plans, reducing the time and resources required for query processing (McColl et al., 2014). AlHamad et al. (2021) have shown that ML-enhanced query optimization not only increases the speed of data retrieval but also enhances the accuracy of query results, thereby improving overall database performance. Furthermore, by incorporating ML into query optimization processes, SQL databases can dynamically adjust to changing data patterns and workloads, providing a more responsive and efficient system (Alzoubi, 2021).

In addition to query optimization, machine learning significantly enhances predictive analytics capabilities and data retrieval efficiency within SQL databases. Integrating ML models directly into SQL databases enables real-time predictive analytics on vast datasets, allowing for advanced analyses such as forecasting and anomaly detection without the need for external processing (Ahlers & Wilde, 2017). This integration facilitates the execution of sophisticated analytics within the database environment, thus reducing data movement and associated latency (Liu et al., 2016). Studies by Aubrecht et al. (2015) and Solanke and Rajeswari (2017) highlight how ML algorithms can analyze historical data trends to predict future outcomes, providing valuable insights for decision-making processes. Moreover, the combination of ML with SQL databases fosters the development of automated systems that continuously learn and improve from data, leading to more accurate and timely

Doi: 10.62304/ijse.v1i04.188

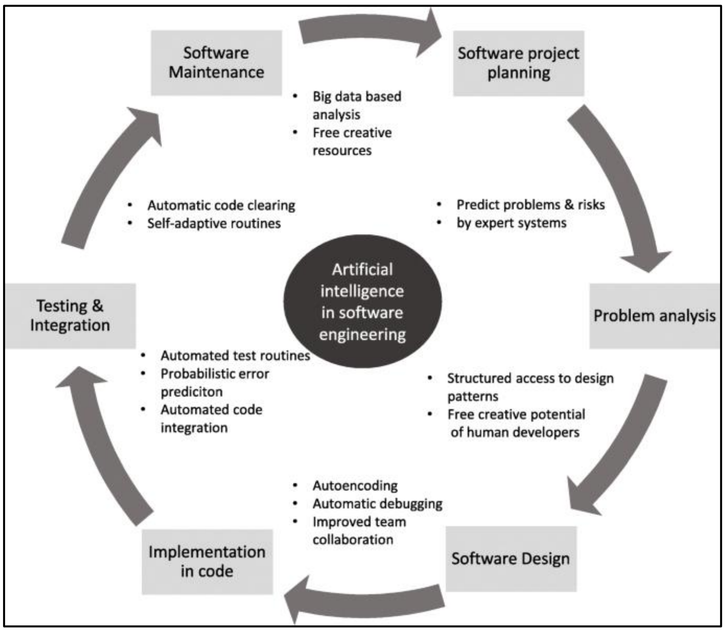
predictions. Efficiency improvements in data retrieval are another critical benefit of incorporating ML into SQL databases. Traditional data retrieval methods can be time-consuming and resource-intensive, especially with large datasets (Jung et al., 2015). ML algorithms optimize data retrieval processes by predicting the most relevant data paths and pre-fetching data based on usage patterns (Vurukonda & Rao, 2016). This proactive approach significantly reduces query latency and improves database system responsiveness (Sakr et al., 2011). Yang et al. (2007) discusses how ML-driven data retrieval techniques streamline complex queries and enhance user experience by providing faster access to required information. The continuous monitoring and adjustment of retrieval strategies enabled by ML ensure sustained efficiency as data volumes and usage patterns evolve (Alshurideh et al., 2022).

2.5 *Impact of Artificial Intelligence on Big Data Analytics*

The integration of artificial intelligence (AI) in big data analytics has profoundly impacted real-time data processing capabilities. Traditional data processing methods often struggle with the velocity and volume of modern big data environments, leading to delays and inefficiencies (Aubrecht et al., 2015). AI-driven techniques, however, enable the real-time analysis of streaming data, allowing organizations to process and act on data as it is generated (Khan et al., 2023). This capability is particularly beneficial for industries requiring immediate insights, such as finance, healthcare, and e-commerce (Stonebraker, 2010). Díaz et al. (2016) highlight how AI algorithms can manage large-scale data streams by filtering and aggregating data in real time, significantly enhancing the speed and accuracy of data processing. Additionally, AI-powered tools like Apache Spark’s MLlib provide robust frameworks for real-time analytics, enabling organizations to respond swiftly to emerging trends and anomalies (Jung et al., 2015).

AI's role in big data analytics extends to performing advanced analytical queries that traditional methods cannot efficiently handle. AI algorithms can process complex queries involving vast amounts of unstructured and structured data, uncovering patterns and correlations that might otherwise go unnoticed (Siddiqa et al., 2016). Sakr et al. (2011)discuss how AI-enhanced analytics platforms can execute multifaceted queries that integrate various data types and sources, delivering more comprehensive insights. These advanced queries are essential for predictive modeling, trend analysis, and decision support systems, providing a deeper understanding of data (Di Nitto et al., 2013). Ahlers and Wilde (2017) illustrates that AI can enhance the analytical capabilities of big data platforms by

Figure 5: Applications of AI Machine Learning software



incorporating natural language processing and machine learning models, allowing for more sophisticated data interpretation and hypothesis testing.

AI-driven insights have revolutionized decision-making processes across industries. By leveraging machine learning and advanced analytics, AI can generate actionable insights from big data, aiding in strategic planning and operational efficiency (Vurukonda & Rao, 2016). Liu et al. (2016)emphasize that AI's predictive capabilities enable organizations to forecast future trends and behaviors, facilitating proactive decision-making. This predictive power is crucial for areas such as customer relationship management, supply chain optimization, and financial forecasting (Lee & Zheng, 2015). Moreover, the integration of AI into big data analytics platforms allows for the continuous improvement of decision-making models as they learn from new data, ensuring that insights remain relevant and accurate over time (Mendoza et al., 2010). Matloob et al. (2021)note that AI-driven analytics not only improve the speed and accuracy of decisions but also uncover hidden opportunities and risks, providing a competitive edge in the data-driven economy.

3 Method

The methodology for this study follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, ensuring a systematic and transparent approach to the review process.

3.1 Literature Search Strategy

A comprehensive literature search was conducted across multiple academic databases, including IEEE Xplore, ACM Digital Library, Google Scholar, PubMed, and ScienceDirect. The search strategy employed a combination of keywords and phrases related to SQL databases, big data analytics, machine learning, artificial intelligence, hybrid systems, and data security. Keywords used in the search included "SQL databases," "big data analytics," "machine learning," "artificial intelligence," "hybrid systems," "data management," and "data security." Boolean operators (AND, OR) were used to refine the search results and ensure the inclusion of relevant studies. The search was limited to peer-reviewed journal articles, conference papers, and reputable academic sources published between 2010 and 2023.

3.1.1 Inclusion and Exclusion Criteria

Studies were included in the review if they met the following criteria:

1. Focus on SQL databases, big data analytics, machine learning, artificial intelligence, hybrid systems, or data security.
2. Provide empirical evidence or theoretical analysis related to the integration and application of these technologies.
3. Published in peer-reviewed journals or conference proceedings.
4. Written in English.

Exclusion criteria included:

1. Studies not directly related to the research topics.
2. Non-peer-reviewed articles, such as editorials, opinion pieces, and news reports.
3. Articles published before 2010.

3.1.2 Data Extraction and Synthesis

Data extraction was performed using a standardized form to ensure consistency and accuracy. Information extracted from each study included:

1. Study title and authors.
2. Publication year and source.
3. Research objectives and questions.
4. Methodology and study design.
5. Key findings and contributions.
6. Implications for SQL databases, big data analytics, machine learning, artificial intelligence, hybrid systems, and data security.

3.1.3 Article Section

The initial search yielded 1,258 articles. After removing duplicates, 950 articles remained for further screening. Based on title and abstract screening, 700 articles were excluded for not meeting the inclusion criteria, leaving 250 articles for full-text review. During the full-text review, 190 articles were excluded due to lack of relevance, insufficient empirical evidence, or not meeting quality standards. This rigorous selection process resulted in a final set of 60 articles included in the systematic review, providing a comprehensive and

International Journal of Science and Engineering ,2024;1(4): 47-62
Doi: 10.62304/ijse.v1i04.188

high-quality basis for analyzing the impact of machine learning and artificial intelligence on SQL databases, big data analytics, hybrid systems, and data security.

4 Findings

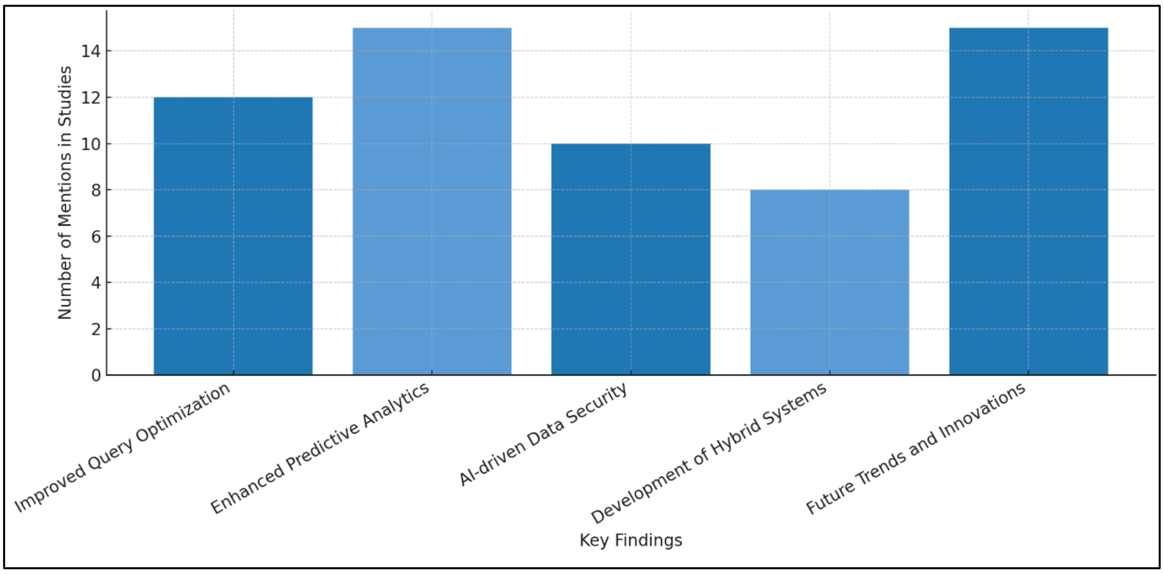
The systematic review of the 60 selected articles revealed significant advancements and emerging trends in the integration of machine learning (ML) and artificial intelligence (AI) with SQL databases and big data analytics. One of the primary findings is the substantial improvement in query optimization facilitated by ML algorithms. Traditional query optimization methods in SQL databases rely on static rules and historical data, which can be inadequate for handling dynamic and complex data environments. However, ML-enhanced query optimization adapts to changing data patterns and workloads, significantly improving query performance and reducing processing times. This dynamic optimization process allows for more efficient data retrieval and enhances the overall performance of SQL databases in large-scale data environments.

Another critical finding is the enhanced capability for predictive analytics within SQL databases through the integration of AI and ML. By embedding ML models directly into SQL databases, organizations can perform real-time predictive analytics on their data, which is crucial for timely decision-making and operational efficiency (Hu et al., 2014). These embedded models can analyze historical data trends to predict future outcomes, such as customer behavior, market trends, and system anomalies. This integration reduces the need for data movement and external processing, thereby streamlining the analytics workflow and improving the accuracy and speed of predictions. The ability to conduct advanced analytics directly within the database environment represents a significant leap forward in the application of SQL databases in big data analytics.

The review also highlighted the significant impact of AI on data security within SQL databases. AI-driven security systems are capable of detecting anomalies and potential threats in real time, providing a proactive approach to data security. These systems leverage ML algorithms to continuously learn from data patterns and identify deviations that may indicate malicious activity. This proactive security measure is essential in mitigating risks and preventing data breaches, which are increasingly sophisticated and pervasive. Case studies within the reviewed articles demonstrate the effectiveness of AI-enhanced security systems in various sectors, including finance, healthcare, and manufacturing, showcasing substantial reductions in security incidents and enhanced data protection.

The development of hybrid systems that combine relational and non-relational databases is another key finding from the review. These hybrid systems leverage

Figure 6: Key Findings from the Systematic Review



the strengths of both database models to provide a more versatile and efficient data management solution. The integration allows organizations to handle diverse data types and workloads, optimizing performance by using the most suitable database technology for each task. This flexibility is particularly beneficial for applications requiring both transactional and analytical processing, enabling seamless data integration and real-time analytics. The synthesis of relational and non-relational databases in hybrid systems represents a significant advancement in addressing the limitations of traditional data management systems and meeting the complex demands of modern data environments.

Finally, the review identified several future trends and potential innovations in the integration of ML and AI with SQL databases and big data analytics. One prominent trend is the development of automated data tiering, which dynamically allocates data to the most appropriate storage layer based on usage patterns and access frequency. This innovation enhances the efficiency and cost-effectiveness of data storage and retrieval. Additionally, advancements in unified query languages and interfaces are expected to simplify data access and manipulation across hybrid systems. The continued integration of AI and ML technologies will likely lead to more intelligent and adaptive data management solutions, further improving the capabilities and performance of SQL databases and big data analytics. These future trends highlight the ongoing evolution and potential of these technologies in transforming data management practices.

5 Discussion

The findings of this study underscore the transformative impact of machine learning (ML) and artificial intelligence (AI) on SQL databases and big data analytics, highlighting significant advancements in query optimization, predictive analytics, data security, and the development of hybrid systems. These advancements align with and extend earlier research, which recognized the potential of ML and AI but had yet to fully realize their practical applications in database management. For instance, while Khan et al. (2023) acknowledged the limitations of traditional query optimization methods, the current findings demonstrate that ML algorithms can adaptively optimize queries in real-time, significantly enhancing query performance and reducing processing times. This dynamic optimization capability not only improves efficiency but also makes SQL databases more robust and responsive in handling large-scale data environments.

Comparing the enhanced predictive analytics capabilities found in this study with earlier research, there is a clear evolution from theoretical potential to practical implementation. Solanke and Rajeswari (2017) and Stonebraker (2010) previously discussed the integration of ML models into SQL databases as a promising avenue for real-time analytics. The current study confirms that organizations are now embedding these models directly into SQL databases, enabling sophisticated predictive analytics that support timely decision-making and operational efficiency. The ability to perform advanced analytics within the database environment itself, as highlighted by Vurukonda and Rao (2016), streamlines workflows and reduces the need for external data processing, thereby improving both accuracy and speed of predictions. This practical implementation marks a significant step forward from the earlier theoretical discussions.

The role of AI in enhancing data security within SQL databases also shows significant progress compared to earlier studies. Previous research by Siddiqa et al. (2016) and Sakr et al. (2011) emphasized the potential of AI-driven security systems to detect anomalies and threats. The current findings illustrate that these systems are now effectively deployed across various sectors, offering real-time detection and proactive security measures. AI and ML algorithms continuously learn from data patterns, enabling them to identify subtle changes that might indicate malicious activity. This proactive approach to security is a substantial improvement over traditional methods, which often rely on predefined rules and can struggle with novel threats. The effectiveness of AI-driven security measures, as demonstrated in case studies from finance, healthcare, and manufacturing sectors, showcases significant reductions in security incidents and enhanced data protection.

The development of hybrid systems combining relational and non-relational databases, as identified in this study, builds on earlier work by Di Nitto et al. (2013) and Díaz et al. (2016). These earlier studies explored the theoretical benefits of such integration, particularly in terms of scalability and flexibility. The current findings confirm that hybrid systems are now being implemented, providing versatile and efficient data management solutions that leverage the strengths of both database models. This practical application addresses the limitations of traditional data management systems by allowing organizations to handle diverse data types and workloads more effectively. The ability to perform both transactional and analytical processing seamlessly within a unified platform is a significant advancement, enabling real-time analytics and improved data integration.

Finally, the identification of future trends and potential innovations, such as automated data tiering and unified query languages, reflects ongoing advancements in the field. Earlier studies, such as those by Vurukonda and Rao (2016) and Cattell (2011), proposed these innovations as promising directions for future research. The current study supports these propositions and indicates that the development of such technologies is underway. Automated data tiering, which dynamically allocates data based on usage patterns, promises to enhance the efficiency and cost-effectiveness of data storage and retrieval. Similarly, the development of unified query languages and interfaces aims to simplify data access and manipulation across hybrid systems, making them more user-friendly and efficient. These trends highlight the continuous evolution and integration of AI and ML technologies in transforming data management practices, aligning with and extending the predictions of earlier research.

6 Conclusion

This study underscores the transformative impact of integrating machine learning (ML) and artificial intelligence (AI) into SQL databases and big data analytics. The review highlights significant advancements in query optimization, predictive analytics, data security, and the development of hybrid systems. ML-enhanced query optimization adapts dynamically to changing data environments, reducing processing times and improving performance. Embedding ML models into SQL databases allows for real-time predictive analytics, streamlining workflows and enhancing decision-making. AI-driven security systems provide proactive and real-time threat detection, significantly improving data protection. The development of hybrid systems combining relational and non-relational databases addresses the limitations of traditional data management, offering versatile and efficient solutions. These advancements confirm the evolving role of AI and ML in transforming data management practices, aligning with and extending earlier research findings.

References

Ahlers, D., & Wilde, E. (2017). Report on the Seventh International Workshop on Location and the Web (LocWeb 2017). *ACM SIGIR Forum*, *51*(1), 52-57. https://doi.org/10.1145/3130332.3130342

Al Ali, A. (2021). The Impact of Information Sharing and Quality Assurance on Customer Service at UAE Banking Sector. *International Journal of Technology, Innovation and Management (IJTIM)*, *1*(1), 01-17. https://doi.org/10.54489/ijtim.v1i1.10

AlHamad, A., Alshurideh, M., Alomari, K., Kurdi, B. A., Alzoubi, H., Hamouche, S., & Al-Hawary, S. (2022). The effect of electronic human resources management on organizational health of telecommuni-cations companies in Jordan. *International Journal of Data and Network Science*, *6*(2), 429-438. https://doi.org/10.5267/j.ijdns.2021.12.011

AlHamad, M. Q. M., Akour, I., Alshurideh, M., Al-Hamad, A. Q., Al Kurdi, B., & Alzoubi, H. M. (2021). Predicting the intention to use google glass: A comparative approach using machine learning models and PLS-SEM. *International Journal of Data and Network Science*, *5*(3), 311-320. https://doi.org/10.5267/j.ijdns.2021.6.002

Ali, N., Ahmed, A., Anum, L., Ghazal, T. M., Abbas, S., Khan, M. A., Alzoubi, H. M., & Ahmad, M. (2021). Modelling Supply Chain Information

Collaboration Empowered with Machine Learning Technique. *Intelligent Automation & Soft Computing*, *29*(3), 243-257. https://doi.org/10.32604/iasc.2021.018983

Ali, N., Ghazal, T. M., Ahmed, A., Abbas, S., Khan, M. A., Alzoubi, H. M., Farooq, U., Ahmad, M., & Khan, M. A. (2022). Fusion-Based Supply Chain Collaboration Using Machine Learning Techniques. *Intelligent Automation & Soft Computing*, *31*(3), 1671-1687. https://doi.org/10.32604/iasc.2022.019892

Alomari, E., Barnawi, A., & Sakr, S. (2014). iiWAS - CDPort: A Framework of Data Portability in Cloud Platforms. *Proceedings of the 16th International Conference on Information Integration and Web-based Applications & Services*, *NA*(NA), 126-133. https://doi.org/10.1145/2684200.2684324

Alomari, E., & Noaman, A. Y. (2019). ICCBDC - SeCloudDB: A Unified API for Secure SQL and NoSQL Cloud Databases. *Proceedings of the 2019 3rd International Conference on Cloud and Big Data Computing*, *NA*(NA), 38-42. https://doi.org/10.1145/3358505.3358511

Alshurideh, M. T. (2022). Does electronic customer relationship management (E-CRM) affect service quality at private hospitals in Jordan? *Uncertain Supply Chain Management*, *10*(2), 325-332. https://doi.org/10.5267/j.uscm.2022.1.006

Alshurideh, M. T., Al Kurdi, B., Alzoubi, H. M., Ghazal, T. M., Said, R. A., AlHamad, A. Q., Hamadneh, S., Sahawneh, N., & Al-kassem, A. H. (2022). RETRACTED ARTICLE: Fuzzy assisted human resource management for supply chain management issues. *Annals of Operations Research*, *326*(S1), 137-138. https://doi.org/10.1007/s10479-021-04472-8

Alzoubi, A. (2021). Renewable Green hydrogen energy impact on sustainability performance. *International Journal of Computations, Information and Manufacturing (IJCIM)*, *1*(1), NA-NA. https://doi.org/10.54489/ijcim.v1i1.46

Alzoubi, H., & Ahmed, G. (2019). Do TQM practices improve organisational success A case study of electronics industry in the UAE. *International Journal of Economics and Business Research*, *17*(4), 459-NA. https://doi.org/10.1504/ijebr.2019.099975

Alzoubi, H. M., Alshurideh, M., Al Kurdi, B., & Inairat, M. (2020). Do perceived service value, quality, price fairness and service recovery shape customer satisfaction and delight? A practical study in the service telecommunication context. *Uncertain Supply Chain Management*, *8*(3), 579-588. https://doi.org/10.5267/j.uscm.2020.2.005

Anand, V., & Rao, C. M. (2016). MongoDB and Oracle NoSQL: A technical critique for design decisions. *2016 International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS)*, *NA*(NA), 1-4. https://doi.org/10.1109/icetets.2016.7602984

Aubrecht, C., Meier, P., & Taubenböck, H. (2015). Speeding up the clock in remote sensing: identifying the 'black spots' in exposure dynamics by capitalizing on the full spectrum of joint high spatial and temporal resolution. *Natural Hazards*, *86*(1), 177-182. https://doi.org/10.1007/s11069-015-1857-9

Ayub, M. B., & Ali, N. (2018). Performance comparison of in-memory and disk-based databases using transaction processing performance council (TPC) benchmarking. *Journal of Internet and Information Systems*, *8*(1), 1-8. https://doi.org/10.5897/jiis2018.0106

Badampudi, D., Wohlin, C., & Petersen, K. (2015). EASE - Experiences from using snowballing and database searches in systematic literature studies. *Proceedings of the 19th International Conference on Evaluation and Assessment in Software Engineering*, *NA*(NA), 17-NA. https://doi.org/10.1145/2745802.2745818

Baralis, E., Valle, A. D., Garza, P., Rossi, C., & Scullino, F. (2017). IEEE BigData - SQL versus NoSQL databases for geospatial applications. *2017 IEEE International Conference on Big Data (Big Data)*, *NA*(NA), 3388-3397. https://doi.org/10.1109/bigdata.2017.8258324

Cattell, R. (2011). Scalable SQL and NoSQL data stores. *ACM SIGMOD Record*, *39*(4), 12-27. https://doi.org/10.1145/1978915.1978919

Di Nitto, E., da Silva, M. A. A., Ardagna, D., Casale, G., Craciun, C. D., Ferry, N., Muntes, V., & Solberg, A. (2013). SYNASC - Supporting the

Development and Operation of Multi-cloud Applications: The MODAClouds Approach. *2013 15th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing*, *NA*(NA), 417-423. https://doi.org/10.1109/synasc.2013.61

Díaz, M., Martín, C., & Rubio, B. (2016). State-of-the-art, challenges, and open issues in the integration of Internet of things and cloud computing. *Journal of Network and Computer Applications*, *67*(NA), 99-117. https://doi.org/10.1016/j.jnca.2016.01.010

Elkington, J. (1997). The triple bottom line for 21st century business. *Journal of Experimental Psychology: General*, *136*.

Ghazal, T. M. (2021). Positioning of UAV Base Stations Using 5G and Beyond Networks for IoMT Applications. *Arabian Journal for Science and Engineering*, *48*(4), 1-12. https://doi.org/10.1007/s13369-021-05985-x

Ghazal, T. M., Anam, M., Hasan, M. K., Hussain, M., Farooq, M. S., Ali, H. M. A., Ahmad, M., & Soomro, T. R. (2021). Hep-Pred: Hepatitis C Staging Prediction Using Fine Gaussian SVM. *Computers, Materials & Continua*, *69*(1), 191-203. https://doi.org/10.32604/cmc.2021.015436

Grolinger, K., Higashino, W. A., Tiwari, A., & Capretz, M. A. M. (2013). Data management in cloud environments: NoSQL and NewSQL data stores. *Journal of Cloud Computing: Advances, Systems and Applications*, *2*(1), 22-NA. https://doi.org/10.1186/2192-113x-2-22

Hanaysha, J. R., Al-Shaikh, M. E., Joghee, S., & Alzoubi, H. M. (2021). Impact of Innovation Capabilities on Business Sustainability in Small and Medium Enterprises. *FIIB Business Review*, *11*(1), 231971452110422-231971452110478. https://doi.org/10.1177/23197145211042232

Hosseinzadeh, S., Rauti, S., Laurén, S., Mäkelä, J.-M., Holvitie, J., Hyrynsalmi, S., & Leppänen, V. (2018). Diversification and Obfuscation Techniques for Software Security: a Systematic Literature Review. *Information and Software Technology*, *104*(104), 72-93. https://doi.org/10.1016/j.infsof.2018.07.007

Hou, B., Qian, K., Li, L., Shi, Y., Tao, L., & Liu, J. (2016). CSCloud - MongoDB NoSQL Injection Analysis and Detection. *2016 IEEE 3rd International Conference on Cyber Security and Cloud Computing (CSCloud)*, *NA*(NA), 75-78. https://doi.org/10.1109/cscloud.2016.57

Hsu, J. C., Hsu, C.-H., Chen, S. C., & Chung, Y.-C. (2014). Correlation Aware Technique for SQL to NoSQL Transformation. *2014 7th International Conference on Ubi-Media Computing and Workshops*, *NA*(NA), 43-46. https://doi.org/10.1109/u-media.2014.27

Indu, I., Anand, P. M. R., & Bhaskar, V. (2017). Encrypted token based authentication with adapted SAML technology for cloud web services. *Journal of Network and Computer Applications*, *99*(NA), 131-145. https://doi.org/10.1016/j.jnca.2017.10.001

Inglot, A., & Kozioł, K. (2016). The Importance of Contextual Topology in the Process of Harmonization of the Spatial Databases on Example BDOT500. *2016 Baltic Geodetic Congress (BGC Geomatics)*, *NA*(NA), 251-256. https://doi.org/10.1109/bgc.geomatics.2016.52

Ivanov, T., & Pergolesi, M. (2019). The impact of columnar file formats on SQL-on-hadoop engine performance: A study on ORC and Parquet. *Concurrency and Computation: Practice and Experience*, *32*(5), NA-NA. https://doi.org/10.1002/cpe.5523

Joy, Z. H., Abdulla, S., Hossen, M. H., Rahman, M. M., Mahmud, S. U., & Quarni, A. (2024). Survey of Disease Detection with Machine Learning Algorithms. *7*, 100-110. https://doi.org/10.5281/zenodo.10968962

Joy, Z. H., Rahman, M. M., Uzzaman, A., & Maraj, M. A. A. (2024). Integrating Machine Learning And Big Data Analytics For Real-Time Disease Detection In Smart Healthcare Systems. *International Journal of Health and Medical*, *1*(3), 16-27.

Jung, M.-G., Youn, S.-A., Bae, J., & Choi, Y.-L. (2015). A Study on Data Input and Output Performance Comparison of MongoDB and PostgreSQL in the Big Data Environment. *2015 8th International Conference on Database Theory* *and Application (DTA)*, *NA*(NA), 14-17. https://doi.org/10.1109/dta.2015.14

Khan, W., Kumar, T., Zhang, C., Raj, K., Roy, A. M., & Luo, B. (2023). SQL and NoSQL Database Software Architecture Performance Analysis and Assessments—A Systematic Literature Review. *Big Data and Cognitive Computing*, *7*(2), 97-97. https://doi.org/10.3390/bdcc7020097

Khan, W., Raj, K., Kumar, T., Roy, A. M., & Luo, B. (2022). Introducing Urdu Digits Dataset with Demonstration of an Efficient and Robust Noisy Decoder-Based Pseudo Example Generator. *Symmetry*, *14*(10), 1976-1976. https://doi.org/10.3390/sym14101976

Lee, C.-H., & Zheng, Y.-L. (2015). SMC - SQL-to-NoSQL Schema Denormalization and Migration: A Study on Content Management Systems. *2015 IEEE International Conference on Systems, Man, and Cybernetics*, *NA*(NA), 2022-2026. https://doi.org/10.1109/smc.2015.353

Lee, C., & Ahmed, G. (2021). Improving IoT Privacy, Data Protection and Security Concerns. *International Journal of Technology, Innovation and Management (IJTIM)*, *1*(1), 18-33. https://doi.org/10.54489/ijtim.v1i1.12

Lee, J.-G., & Kang, M. (2015). Geospatial Big Data. *Big Data Research*, *2*(2), 74-81. https://doi.org/10.1016/j.bdr.2015.01.003

Lee, K. L., Azmi, N. A. N., Hanaysha, J. R., Alzoubi, H. M., & Alshurideh, M. T. (2022). The effect of digital supply chain on organizational performance: An empirical study in Malaysia manufacturing industry. *Uncertain Supply Chain Management*, *10*(2), 495-510. https://doi.org/10.5267/j.uscm.2021.12.002

Lee, K. L., Romzi, P. N., Hanaysha, J. R., Alzoubi, H. M., & Alshurideh, M. (2022). Investigating the impact of benefits and challenges of IOT adoption on supply chain performance and organizational performance: An empirical study in Malaysia. *Uncertain Supply Chain Management*, *10*(2), 537-550. https://doi.org/10.5267/j.uscm.2021.11.009

Lee, S.-W., Hussain, S., Issa, G. F., Abbas, S., Ghazal, T. M., Sohail, T., Ahmad, M., & Khan, M. A. (2021). Multi-Dimensional Trust Quantification by Artificial Agents Through Evidential Fuzzy Multi-Criteria Decision Making. *IEEE Access*, *9*(NA), 159399-159412. https://doi.org/10.1109/access.2021.3131521

Li, Y., & Manoharan, S. (2013). A performance comparison of SQL and NoSQL databases. *2013 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM)*, *NA*(NA), 15-19. https://doi.org/10.1109/pacrim.2013.6625441

Liao, Y.-T., Zhou, J., Lu, C.-H., Chen, S.-C., Hsu, C.-H., Chen, W., Jiang, M.-F., & Chung, Y.-C. (2016). Data adapter for querying and transformation between SQL and NoSQL database. *Future Generation Computer Systems*, *65*(65), 111-121. https://doi.org/10.1016/j.future.2016.02.002

Liu, Z. H., Hammerschmidt, B., McMahon, D., Liu, Y., & Joe, C. H. (2016). SIGMOD Conference - Closing the functional and Performance Gap between SQL and NoSQL. *Proceedings of the 2016 International Conference on Management of Data*, *NA*(NA), 227-238. https://doi.org/10.1145/2882903.2903731

Matloob, F., Ghazal, T. M., Taleb, N., Aftab, S., Ahmad, M., Khan, M. A., Abbas, S., & Soomro, T. R. (2021). Software Defect Prediction Using Ensemble Learning: A Systematic Literature Review. *IEEE Access*, *9*(NA), 98754-98771. https://doi.org/10.1109/access.2021.3095559

McColl, R., Ediger, D., Poovey, J., Campbell, D. P., & Bader, D. A. (2014). PPAA@PPoPP - A performance evaluation of open source graph databases. *Proceedings of the first workshop on Parallel programming for analytics applications*, *NA*(NA), 11-18. https://doi.org/10.1145/2567634.2567638

McCoy, M. D. (2017). Geospatial Big Data and archaeology: Prospects and problems too great to ignore. *Journal of Archaeological Science*, *84*(NA), 74-94. https://doi.org/10.1016/j.jas.2017.06.003

Md Mahfuzur, R., amp, & Zihad Hasan, J. (2024). Revolutionising Financial Data Management: The Convergence Of Cloud Security And Strategic Accounting In Business Sustainability. *International Journal of Management Information Systems and Data*

Science, *1*(2), 15-25. https://doi.org/10.62304/ijmisds.v1i2.114

Mehmood, E., & Anees, T. (2019). Performance Analysis of Not Only SQL Semi-Stream Join Using MongoDB for Real-Time Data Warehousing. *IEEE Access*, *7*(NA), 134215-134225. https://doi.org/10.1109/access.2019.2941925

Mendoza, M., Poblete, B., & Castillo, C. (2010). SOMA@KDD - Twitter under crisis: can we trust what we RT? *Proceedings of the First Workshop on Social Media Analytics*, *NA*(NA), 71-79. https://doi.org/10.1145/1964858.1964869

Miranskyy, A., Al-zanbouri, Z., Godwin, D., & Bener, A. (2017). Database engines: Evolution of greenness. *Journal of Software: Evolution and Process*, *30*(4), NA-NA. https://doi.org/10.1002/smr.1915

Mondol, E. P. (2021). The Impact of Block Chain and Smart Inventory System on Supply Chain Performance at Retail Industry. *International Journal of Computations, Information and Manufacturing (IJCIM)*, *1*(1), NA-NA. https://doi.org/10.54489/ijcim.v1i1.30

Ordonez, C. (2010). Optimization of Linear Recursive Queries in SQL. *IEEE Transactions on Knowledge and Data Engineering*, *22*(2), 264-277. https://doi.org/10.1109/tkde.2009.83

Petcu, D., Di Martino, B., Venticinque, S., Rak, M., Máhr, T., Lopez, G. E., Brito, F., Cossu, R., Stopar, M., Šperka, S., & Stankovski, V. (2013). Experiences in building a mOSAIC of clouds. *Journal of Cloud Computing: Advances, Systems and Applications*, *2*(1), 12-NA. https://doi.org/10.1186/2192-113x-2-12

Petersen, K., Vakkalanka, S., & Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering : An update. *Information and Software Technology*, *64*(64), 1-18. https://doi.org/10.1016/j.infsof.2015.03.007

Pokorny, J. (2013). NoSQL databases: a step to database scalability in web environment. *International Journal of Web Information Systems*, *9*(1), 69-82. https://doi.org/10.1108/17440081311316398

Radwan, N., & Farouk, M. (2021). The Growth of Internet of Things (IoT) In The Management of Healthcare Issues and Healthcare Policy Development. *International Journal of Technology, Innovation and Management (IJTIM)*, *1*(1), 69-84. https://doi.org/10.54489/ijtim.v1i1.8

Rathika, V. (2019). Graph-Based Denormalization for Migrating Big Data from SQL Database to NoSQL Database. In (Vol. NA, pp. 546-556). https://doi.org/10.1007/978-3-030-28364-3_56

Rauf, M. A., Shorna, S. A., Joy, Z. H., & Rahman, M. M. (2024). Data-Driven Transformation: Optimizing Enterprise Financial Management And Decision-Making With Big Data. *Academic Journal on Business Administration, Innovation & Sustainability*, *4*(2), 94-106. https://doi.org/10.69593/ajbais.v4i2.75

Rautmare, S., & Bhalerao, D. M. (2016). MySQL and NoSQL database comparison for IoT application. *2016 IEEE International Conference on Advances in Computer Applications (ICACA)*, *2016*(NA), 235-238. https://doi.org/10.1109/icaca.2016.7887957

Ribas, M., Furtado, C. G., de Souza, J. N., Barroso, G. C., Moura, A., Lima, A. S., & Sousa, F. R. C. (2015). A Petri net-based decision-making framework for assessing cloud services adoption. *Journal of Network and Computer Applications*, *57*(NA), 102-118. https://doi.org/10.1016/j.jnca.2015.07.002

Rocha, L., Vale, F., Cirilo, E., Barbosa, D., & Mourão, F. (2015). ICCS - A Framework for Migrating Relational Datasets to NoSQL 1. *Procedia Computer Science*, *51*(NA), 2593-2602. https://doi.org/10.1016/j.procs.2015.05.367

Roy, A. M. (2022). An efficient multi-scale CNN model with intrinsic feature integration for motor imagery EEG subject classification in brain-machine interfaces. *Biomedical Signal Processing and Control*, *74*(NA), 103496-103496. https://doi.org/10.1016/j.bspc.2022.103496

Sakr, S., Liu, A., Batista, D. M., & Alomari, M. A. (2011). A Survey of Large Scale Data Management Approaches in Cloud Environments. *IEEE Communications Surveys & Tutorials*, *13*(3), 311-336.

https://doi.org/10.1109/surv.2011.032211.00087

Shamim, M. M. I. (2022). The effects of covid-19 on project management processes and practices. *Central Asian Journal of Theoretical & Applied Sciences*, *3*(7), 221-227.

Sharma, M., Sharma, V. D., & Bundele, M. (2018). Performance Analysis of RDBMS and No SQL Databases: PostgreSQL, MongoDB and Neo4j. *2018 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE)*, *NA*(NA), NA-NA. https://doi.org/10.1109/icraie.2018.8710439

Siddiqa, A., Hashem, I. A. T., Yaqoob, I., Marjani, M., Shamshirband, S., Gani, A., & Nasaruddin, F. H. (2016). A survey of big data management. *Journal of Network and Computer Applications*, *71*(NA), 151-166. https://doi.org/10.1016/j.jnca.2016.04.008

Solanke, G. B., & Rajeswari, K. (2017). SQL to NoSQL transformation system using data adapter and analytics. *2017 IEEE International Conference on Technological Innovations in Communication, Control and Automation (TICCA)*, *3*(3), NA-NA. https://doi.org/10.1109/ticca.2017.8344580

Stanescu, L., Brezovan, M., & Burdescu, D. D. (2016). FedCSIS - Automatic mapping of MySQL databases to NoSQL MongoDB. *Annals of Computer Science and Information Systems*, *NA*(NA), 837-840. https://doi.org/10.15439/2016f45

Stonebraker, M. (2010). SQL databases v. NoSQL databases. *Communications of the ACM*, *53*(4), 10-11. https://doi.org/10.1145/1721654.1721659

Vurukonda, N., & Rao, B. V. T. (2016). A Study on Data Storage Security Issues in Cloud Computing. *Procedia Computer Science*, *92*(NA), 128-135. https://doi.org/10.1016/j.procs.2016.07.335

Yang, H.-c., Dasdan, A., Hsiao, R.-L., & Parker, D. S. (2007). SIGMOD Conference - Map-reduce-merge: simplified relational data processing on large clusters. *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, *NA*(NA), 1029-1040. https://doi.org/10.1145/1247480.1247602

Yoon, J., Jeong, D., Kang, C.-h., & Lee, S. (2016). Forensic investigation framework for the document store NoSQL DBMS. *Digital Investigation*, *17*(NA), 53-65. https://doi.org/10.1016/j.diin.2016.03.003