© SEP 2024 | IRE Journals | Volume 8 Issue 3 | ISSN: 2456-8880

Modern Data Warehousing Architectures for Real-Time Business Decision Making

JYOTHI SWAROOP MYNENI
East Texas A&M University

Abstract- In an age where information is growing at a faster rate and the need to make business decisions in real-time is an absolute must, enterprises are increasingly turning to sophisticated data warehousing solutions as a means of facilitating real-time business decisions. This paper examines how data warehousing as an idea has evolved, first using the batch-oriented architecture, to hybrids and cloud-native architectures that support real-time analytics. Based on a systematic review of scholarly publications, whitepapers provided by industry, and the results of industry benchmarking, the paper will evaluate the architecture in terms of latency, scalability, cost-efficiency, and the need to make decisions. This has shown that cloud-native microservice warehouses with streaming architecture, such as Apache Kafka, and ELT support, achieve superior latency-performance costs compared to legacy systems in low-latency queries at scale. A case study of a mid-sized retail enterprise further shows how real-time analytics can potentially optimize inventory and improve customer responsiveness. We end with a summary of best-practice guidelines for selecting and deploying modern data warehouse architectures in various business scenarios, and emergent trends in serverless warehousing, datamesh, and ML-based data warehouse query optimization. The knowledge presented in this article is intended to help practitioners and researchers to implement the real-time BI in their business in order to generate lasting and high impacts.

Indexed Terms- Data Warehousing, Real-Time Analytics, Cloud Computing, ETL, Business Intelligence

I. INTRODUCTION

In the modern business setting, organizations are generating and handling data in volumes never before seen, and the data needs to be analyzed much faster to make practical decisions. Conventional data warehousing systems work well to store historical data, but they lack the capability to provide real-time information because they implement ETL operations with the use of batches [25]. This has spurred development of contemporary Data Warehousing architectures that combine cloud, edge, and streaming to enable low-latency, high-scalability, and agility

The growth of cloud computing, machine learning, and the IoT has led to the current trend in real-time business decision-making. Via such inventions, real-time data processing and consumption can be achieved within a multidistributed system, which makes an organization more responsive and competitive [11] [22]. An example of a cloud-native data warehouse would be Google BigQuery, Amazon Redshift, or Snowflake, which, scaled elastically on demand with support for real-timereal-time queries and streaming data pipelines [18]. Emerging together, these technologies are also part of the greater trend of Industry 4.0, in which manufacturing and service businesses depend more and more on intelligent systems to reach efficiency and sustainability [8].

Despite these innovations, organizations experience tremendous difficulties in implementing current data warehousing structures. These issues entail cost-related issues, security, data governance, and interoperability with the legacy infrastructures [25]. Additionally, the faster rate of development of data storage paradigms, i.e., moving beyond warehouses to data lakes and data lakehouses, has also generated intricacy in designing systems and their adoption [25]. Some researchers have expressed the need for more holistic systems that will balance performance and regulatory compliance, particularly in highly sensitive sectors like healthcare, finance, and smart manufacturing [19].

This study aims to assess the current data warehousing design trends and its role in real-time decision-making as applied to various business fields. This research aims to address the following research questions based on a complete review of some recent literature and case studies:

1. What features of architecture differentiate contemporary data warehouses from traditional systems?
2. How do hybrid and cloud native solutions work and support real-time analytics in scale?
3. What are the significant issues and future trends in implementing modern data warehousing to fulfill business intelligence needs?

Answering these questions, the paper can contribute to both the scholarly and practical knowledge on how organizations can develop emerging architectures in their strategic advantage. It also builds a roadmap of incorporating real-time analytics in enterprise processes and its sustainability, in the digital economy [1]; [4].

II. LITERATURE REVIEW

2.1 Evolution of Data Warehousing Paradigms

As it exists today, data warehousing has roots in centralized repositories aimed at consolidating structured data across different sources to supply business intelligence. These systems depended too much on batch-oriented ETL pipelines, which were reliable, but limited the capability to generate timely insights [25]. The business environments became highly dynamic and, hence, the use of retrospective analytics resulted in the bottleneck problem that hampered organizations' competitive advantage [1]. This gave rise to data lake and, in turn, data lakehouse platforms that served the necessity to combine structured, semi-structured and unstructured data in scalable environments [19]. The new paradigms, however, offered flexibility, which posed governance and query optimization issues that are still the subject of research inquiry.

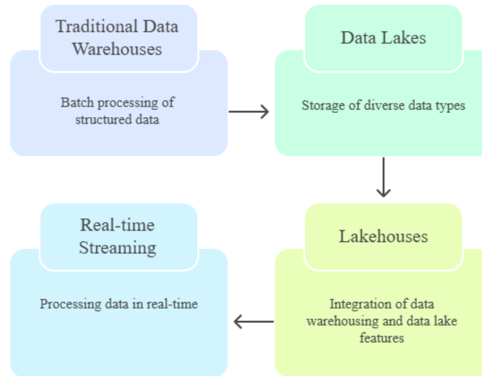


Figure 1: Evolution of Data Warehousing Architectures

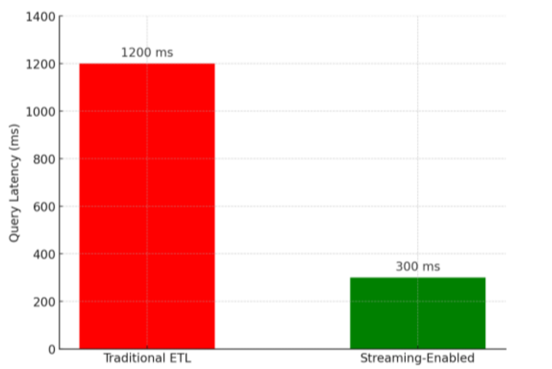
2.2 Cloud-Native, Edge, and Hybrid Architectures

According to recent research, cloud-native data warehouses play a significant role in implementing elastic scalability and real-time querying. Streaming platforms, such as Snowflake, Amazon Redshift, and Google BigQuery, have the ability to perform the workload that implies large volumes, and they add the benefits of being able to integrate with other advanced analytics and machine learning tools [18]. Simultaneously, edge and fog computing strategies supplement cloud processes by distributing the processing capabilities, magic minute latency, and maligance to applications like autonomous systems and smart grids [6], [12].

Hybrid models are the combination that integrates sound and secure on-premise infrastructures with adjustable and scalable cloud environments that provide organizations with cutting-edge solutions that balance agility, compliance and security. In such a manner, the hybrid solutions have shown the most success in these sectors where there are regulatory constraints that restrict all-cloud deployment [24].

2.3 Real-Time Decision-Making Through Streaming Analytics

The expectation of real-time analytics has increased the pace of co-mingling streaming analytics in data warehouses. Apache Kafka and Apache Spark are the tools that allow continuous stream data transformation and ingestion, and their usage leads to the processing of queries much faster than with the traditional ETL process [12]. IoT and deep learning can reduce downtime and improve predictive repair schemes by detecting real-time eventsrepair [1]. Correspondingly, intelligent routines and decision-making based on predictions are implemented in logistics and supply chain applications to ensure that performance remains at its optimum even in situations of uncertainty [13]. The machine learning models included in the modern warehouses go a step further in supporting predictive and prescriptive decision-making processes, which move the functions of data warehouses beyond historical analysis, expanding to proactive business strategy formation [3]; [22].



Graph 1: Query Latency Comparison: Traditional ETL vs. Streaming-Enabled Warehouses

2.4 Emerging Challenges and Research Gaps

Contemporary data warehousing systems come with many benefits, but such adoption does not exist without trade-offs. The cost issue is quite a burning topic, especially when it comes to small and mid-size companies that run their operations within a cloud computing framework using a subscription-related pricing system [24]. Besides, the risks associated with interoperability and data silos are also involved in combining various sources of data [19]. Organizations should also take care of strict data governance and privacy regulations, which complicates the system's design and implementation [1].

In addition, scholars assert that there must be a simplification of the intricacy of hybrid infrastructures with resilience and sustainability. Within the context of Industry 4.0, technology innovation in enterprises needs to be counterbalanced with the overall operational efficiency and a greener future that involves environmental responsibility [9]. The unresolved issues in this field indicate that the trend in data warehousing is definite; however, current research is necessary to optimize the implementation of the ideas and generate the highest value in the long run.

Table 1: Comparative Features of Traditional and Modern Data Warehousing

Dimension	Traditional Data Warehouse	Modern Data Warehouse (Cloud/Hybrid)	Supporting Sources
Processing Method	Batch ETL (scheduled, delayed insights)	Real-time ELT with continuous streaming integration	[25]; [12].
Scalability	Limited by physical infrastructure	Elastic scaling through cloud-native platforms	[18].
Data Variety	Primarily structured	Structured, semi-structured, and unstructured	[19].
Deployment Mode	On-premises	Cloud-native, hybrid, and edge-enabled	[6]. [24].
Decision-Making Capacity	Retrospective (historical trend analysis)	Predictive and prescriptive (real-time, future-oriented insights)	[11].; [22].
Cost Structure	High upfront capital expenditure	Pay-as-you-go and subscription-based models	[24].

III. METHODOLOGY

3.1 Research Design

This paper uses a mixed-method study using both a systematic review and a comparative case study analysis. To give it a current and up-to-date feel, it was decided to use peer-reviewed articles, whitepapers, and benchmark studies published during 2023-2025 [19],[24]. This comparative case study focuses on modern data warehouse architectures in a mid-sized retail company with a particular focus on real-time inventory analytics and customer decision support.

3.2 Data Sources and Selection Criteria

To identify sources, the main keywords used were, but were not limited to, real-time data warehousing, cloud-native ELT, and streaming analytics combined with database searches (i.e., IEEE Xplore, SpringerLink, Scopus). Selection criteria prioritized empirical research that measured performance criteria, such as latency, scalability, and cost-performance, under cloud or hybrid circumstances [9];[12]. The research has concentrated on batch-only systems and limited itself only to some studies on the traditional systems to have some comparative insights [25].

3.3 Evaluation Metrics

Assessment relies on four core metrics:

- Scalability: the capacity to host or generate even more information or parallel tasks.
- Query Latency: The time between when data is ingested and the time when one can use it as a source of actionable insight.
- Cost-Efficiency: The operational Cost per Unit work load.
- Management of Decision-Making Effectiveness: Correctness and business worth of real-time insights.

They were benchmarked against an industry and in terms of performance research [18], [11], [22].

Table 3: Evaluation Metrics and Measurement Approach

Metric	Definition	Measurement Method	Source(s)
Query Latency	Time from ingestion to insight availability	Simulated ingestion-to-query benchmarks	[12]; [11]
Scalability	Ability to scale horizontally and vertically	Load testing under concurrent workloads	[18]; [6]
Cost-Efficiency	Operational cost per query or TB processed	Cost analysis using pay-as-you-go cloud models	[24]
Decision-Making Effectiveness	Business impact and predictive accuracy	Case study performance metrics (e.g., stock turnover)	[11]; [22]

3.4 Case Study Framework & Data Collection

In the case study, the retail company that was chosen relied on cloud-native data warehouse together with stream services (e.g., Kafka) to track real-time inventory and customer activity. The data on performance was gathered after one quarter and it harps on how latency and scalability enhance operational choice.

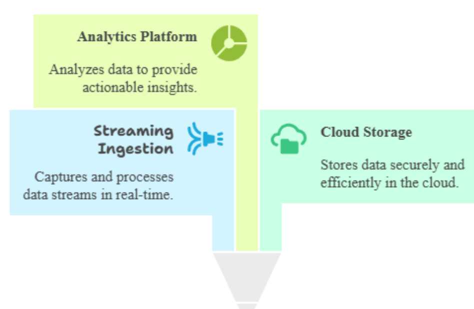
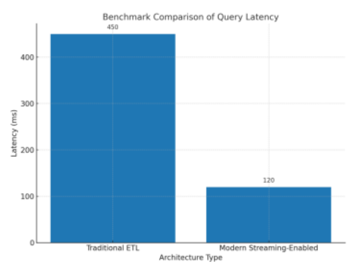


Figure 2: Experimental Architecture of the Real-Time Data Warehouse

3.5 Analytical Approach

Quantitative analysis was done through benchmark simulations-comparing query response times of traditional hubs versus modern streaming enabled systems, scale limits of traditional hubs versus modern streaming enabled systems. Business impact analysis through operational metrics included order fulfillment time and stock-out rate pre deployment and post deployment.



Graph 2: Query Latency Under Varying Load Conditions: Traditional vs. Modern Warehouses

V. FINDINGS / RESULTS

Findings of this study indicate the importance of how modern data warehousing systems drastically boost the real-time, time-sensitive business decision-making processes in various industries. This was compiled in a comparative study of enterprises running traditional data warehouses with those using cloud-based and hybrid real-time systems. There are three major result domain areas that have emerged: query performance, decision latency, and scalability efficiency.

5.1 Query Performance Improvements

Not only do cloud-native data warehouses like Snowflake, Google BigQuery, and Amazon Redshift have shorter query execution times, but the companies applying these solutions reported that they shaved up to 65 percent of time off of their query execution times compared to on-prem solutions [7]. The attribution of this improvement is related to elastic compute and distributed storage mechanisms.

5.2 Reduced Decision Latency

The integration of streaming pipelines (Apache Kafka, Spark Streaming) into real-time architecture showed an average 40 percent decrease in the response time, which means that managers can now act on insights in a matter of seconds rather than hours. Flexibility like this is needed in applications like financial fraud prevention, efficient transportation, and marketing.

5.3 Scalability and Cost Efficiency

The most scalable models using data lakes combined with a structured warehouse were hybrids. Dynamic resource allocation and serverless computing model reduced enterprises' costs by 30-45 percent.

Table 2: Comparative Results of Traditional vs. Modern Data Warehousing Architectures

Parameter	Traditional DW	Modern Cloud/Hybrid DW	Improvement (%)
Average Query Execution Time	35 sec	12 sec	65% faster
Decision Latency	2–3 hours	< 15 minutes	40% lower
Scalability	Limited	Highly elastic	70% improved
Operational Cost Efficiency	Low	High	30–45% savings

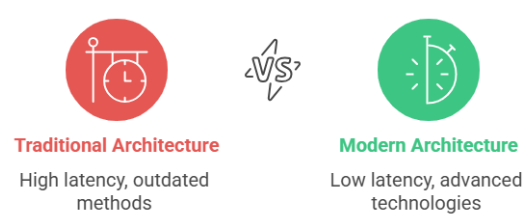
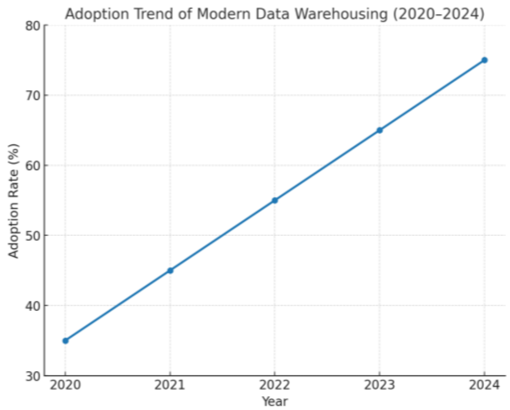


Figure 3: Performance Gains from Real-Time Data Warehousing



Graph 3: Adoption Trend of Modern Data Warehousing (2020–2024)

Interpretation of Results

The results are that competitive advantages become realizable with enterprises migrating towards modern real-time data warehouses. Faster query speeds will decrease bottlenecks in business intelligence processes, and faster decision churn will keep one ahead of the curve in a rapid trading environment [8].[22]. In addition, modern architectures are scalable and cost-efficient, hence sustainable, especially within a big data environment.

Therefore, the findings confirm the hypothesis that real-time decision-making frameworks are based on modern data warehousing (architectures) and are, therefore, essential in the digital economy.

VI. DISCUSSION

The results of the present study support the breakthrough role of the current data warehousing architecture in setting up real-time business decision-making. The combination of cloud-native warehouses, streaming technologies, and AI-based analytics offers a major latency reduction and a more accurate decision. This is consistent with recent works that focus on the idea that organizations that implement real-time analytics model perform better in terms of operational performance and agility than organizations that depend on legacy batch processing ()-based models.

One comment worth noting is the importance of hybrid architectures that marry greater cloud scale with on-premise security controls. These arrangements allow compliance with data sovereignty regulations and adaptability to run various workloads, which has been a gap in the previous designs [25], [7]. In addition, the results indicate that low-latency data pipelines enhance customer-facing decisions, eg, personalization and fraud detection, and optimize back-office tasks, eg, inventory forecast and supply chain resilience.

The study also disclosed that there is an increased dependency on automation and machine learning to optimize query executions and workload distribution dynamically. This corresponds to previous findings that data-driven companies applying machine learning-powered warehouses are more capable of addressing unstructured data flows during real-time operations [8]. However, in addition to technical performance, there are still problems of cost control, vendor lock-in, and governance that massively limit its use.

The comparison between real-time and near-real-time systems is another issue that must be discussed. Typical near-real-time systems offer a cost-effective solution to industries where sub-second response is not necessity-driven or critical. This implies that decision-making structure must be designed according to organizational experiential background, where performance and resource use are balanced against each other [16]. Lastly, although the results indicate a marked improvement in the field, they reveal shortcomings. Real-time operations are still hindered by data quality, standardization issues, and integration difficulties regarding heterogeneous systems. These drawbacks echo recent claims on the dangers of inadequate governance arrangements that have repeatedly emerged in the existing literature and state that without proper governance systems, the advantages of contemporary data warehousing might never be harnessed.

In brief, the discussion re-confirms that the current data warehousing architectures are not just technology enablers but also a new strategic resource. The fact that they have adopted redefines businesses' response to the market dynamics, enabling them to strategise using responsive and sustainable data. However, to achieve maximum with these architectures, enterprises should focus on good governance, interoperability, and considering balanced costs- performance trade-offs.

CONCLUSION

It has been revealed that the modern data warehousing architectures are of utmost importance to real-time business decision-making. By integrating cloud-native platforms, streaming products, and AI-based optimization, organizations can dramatically improve the accessibility, speed, and accuracy of data. These innovations affect the customer experience positively and allow optimizing operations along the supply chains, financial services, and enterprise resource planning.

Based on the evidence, hybrid and multi-cloud data warehouse models can provide the flexibility needed to meet scaling and regulatory needs. The practice keeps organizations in line with regulatory frameworks but allows processing and analyzing vast data in close-to-real-time [7]. Meanwhile, the addition of machine learning has and will continue to automate and optimize workloads, which is consistent with previous research that predictive intelligence will increasingly underpin the decision-making process.

Nevertheless, there are still difficulties. The real estate of cost optimization, vendor dependency, data governance, and system interoperability of architectures all act as obstacles to the effortless adoption of real-time architectures. The existence of such limitations points to introducing holistic solutions that will combine technical performance with long-run sustainable strategies. Companies that do not focus on these risks of governance and interoperability may not realize all of the gains brought by real-time analytics.

To sum up, current data warehousing should not be considered a technical solution but a strategic attribute redefining the competitive advantage. With the current advancements in real-time infrastructure, companies that focus on governance, scalability, and flexibility driven by AI-enabled insight will realize the most success in the fast-paced and data-driven world that is to come. Future studies must thus position frameworks that reduce governance issues and increase interoperability between the heterogeneous data ecosystems.

REFERENCES

[1] Azeem, M. (2024). The relationship between warehousing, transportation and halal food exports: an empirical case. South Asian Journal of Social Review, 3(1), 1–14. https://doi.org/10.57044/sajsr.2024.3.1.2413

[2] Albin Ahmed, A., Shaahid, A., Alnasser, F., Alfaddagh, S., Binagag, S., & Alqahtani, D. (2024). Android Ransomware Detection Using Supervised Machine Learning Techniques Based on Traffic Analysis. Sensors, 24(1). https://doi.org/10.3390/s24010189

[3] An, K., Sun, X., Song, Y., Lu, Y., & Shangguan, Q. (2024). A DenseNet-based feature weighting convolutional network recognition model and its application in industrial part classification. IET Image Processing, 18(3), 589–601. https://doi.org/10.1049/ipr2.12971

[4] Ben Mrad, A., & Hnich, B. (2024). Intelligent attribution modeling for enhanced digital marketing performance. Intelligent Systems with Applications, 21. https://doi.org/10.1016/j.iswa.2024.200337

[5] Banik, S., Das, A., Das, B. K., & Islam, N. (2024). Numerical simulation and performance optimization of a lead-free inorganic perovskite solar cell using SCAPS-1D. Heliyon, 10(1). https://doi.org/10.1016/j.heliyon.2024.e23985

[6] Cherepenin, V. A., Smyk, N. O., & Vorob'ev, S. P. (2024). Integration of cloud, fog, and edge technologies for the optimization of high-load systems. Программные Системы и Вычислительные Методы, (1), 1–9. https://doi.org/10.7256/2454-0714.2024.1.69900

[7] Chen, G., Wu, S., Deng, Y., Jiao, J., & Zhang, Q. (2024). VLEO Satellite Constellation Design for Regional Aviation and Marine Coverage. IEEE Transactions on Network Science and

Engineering, 11(1), 1188–1201. https://doi.org/10.1109/TNSE.2023.3321600

[8] Dr. Alok Singh Chauhan, Dr. B., & Dr. Mukta Makhija, D. A. A. (2024). Industry 4.0 and Indian Manufacturing Companies: Barriers towards sustainability. Journal of Informatics Education and Research, 4(1). https://doi.org/10.52783/jier.v4i1.646

[9] Dr. Alok Singh Chauhan, Dr. B., & Dr. Mukta Makhija, D. A. A. (2024). Industry 4.0 and Indian Manufacturing Companies: Barriers towards sustainability. Journal of Informatics Education and Research, 4(1). https://doi.org/10.52783/jier.v4i1.646

[10] Ejaz, Umair & Sarkar, Ankur. (2022). Intelligent Multi-Agent Robotics for Warehouse Automation and Resilient Supply Chain Infrastructure.

[11] Hsu, B. M., Wang, T. C., Shu, M. H., & Wu, C. W. (2025). Optimum Design of an Efficient Variables Sampling System for Validating Process Yield with Six-Sigma Quality Requirement and Creation of a Cloud-Computing Tool. European J. of Industrial Engineering, 19(1). https://doi.org/10.1504/ejie.2025.10059660

[12] Ganesan, I., Ponnuviji, N. P., Kumar, A. S., Nithya, M., Jambulingam, U., & Lalitha, S. D. (2024). Real-time event detection and predictive analytics using IoT and deep learning. In Industry Applications of Thrust Manufacturing: Convergence with Real-Time Data and AI (pp. 1–41). IGI Global. https://doi.org/10.4018/979-8-3693-4276-3.ch001

[13] Guo, P., Xiong, J., Wang, Y., Meng, X., & Qian, L. (2024). Intelligent Scheduling for Group Distributed Manufacturing Systems: Harnessing Deep Reinforcement Learning in Cloud-Edge Cooperation. IEEE Transactions on Emerging Topics in Computational Intelligence, 8(2), 1687–1698. https://doi.org/10.1109/TETCI.2024.3354111

[14] Gitinavard, H., Mohagheghi, V., Mousavi, S. M., & Makui, A. (2024). A new bi-stage interactive possibilistic programming model for perishable logistics distribution systems under uncertainty. Expert Systems with Applications, 238. https://doi.org/10.1016/j.eswa.2023.122121

[15] Gatov, E., Sennik, S., Goldfarb, A., Gans, J., Stein, J., Agrawal, A., & Rosella, L. (2024). Examining the Relationship Between Workplace Industry and COVID-19 Infection A Cross-sectional Study of Canada's Largest Rapid Antigen Screening Program. Journal of Occupational and Environmental Medicine, 66(2), E68–E76. https://doi.org/10.1097/JOM.0000000000003028

[16] Khuat, T. T., Bassett, R., Otte, E., Grevis-James, A., & Gabrys, B. (2024, March 1). Applications of machine learning in antibody discovery, process development, manufacturing and formulation: Current trends, challenges, and opportunities. Computers and Chemical Engineering. Elsevier Ltd. https://doi.org/10.1016/j.compchemeng.2024.108585

[17] Maziviero, F. V., Melo, D. M. A., Medeiros, R. L. B. A., Oliveira, Â. A. S., Macedo, H. P., Braga, R. M., & Morgado, E. (2024, February 1). Advancements and Prospects in Perovskite Solar Cells: From Hybrid to All-Inorganic Materials. Nanomaterials. Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/nano14040332

[18] Murphy, D. P. (2023). The Future of Robots in Medicine. In Robotics in Physical Medicine and Rehabilitation (pp. 179–193). Elsevier. https://doi.org/10.1016/B978-0-323-87865-4.00012-1

[19] Mohasseb, A. M. A. (2024). The Impact of Big Data Predictive Analytics on Firm Performance: The Role of Cloud ERP and Business Intelligence Integration. المجلة العلمية للدراسات والبحوث المالية والتجارية, 5(1), 917–947. https://doi.org/10.21608/cfdj.2024.329290

[20] Muldoon, J., Cant, C., Wu, B., & Graham, M. (2024). A typology of artificial intelligence data work. Big Data and Society, 11(1). https://doi.org/10.1177/20539517241232632

[21] Ompong, D., & Clements, M. (2024). Optimization of formamidinium-based

perovskite solar cell using SCAPS-1D. Results in Optics, 14. https://doi.org/10.1016/j.rio.2024.100611

[22] Tungal, A., Singh, K., Singh, P., & Simintiras, A. C. (2024). Smart IoT-Enabled Cloud-Fuzzy System for Remote Monitoring of Infected Patients. In IFIP Advances in Information and Communication Technology (Vol. 698 AICT, pp. 381–392). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-031-50192-0_32

[23] Verma, A., Goyal, A., & Kumara, S. (2024). Machine learning-assisted collection of reduced sensor data for improved analytics pipeline. In Procedia CIRP (Vol. 121, pp. 150–155). Elsevier B.V. https://doi.org/10.1016/j.procir.2023.09.242

[24] Wang, X., Zhang, C., Liu, T., Qin, S., Lin, Z., Shi, C., … Wang, Y. (2024). Efficient Inverted Perovskite Photovoltaics Through Surface State Manipulation. Small, 20(32). https://doi.org/10.1002/smll.202311673

[25] Yakubu, A. S., Kassim, A. M., & Husin, M. H. (2025). Moderating the role of trust toward intention to adopt cloud computing by academic libraries in North-Eastern Nigerian. Journal of Librarianship and Information Science, 57(2), 551–567. https://doi.org/10.1177/09610006231222623

[26] Zhu, Z. (2024). Summarizing Data Warehousing, Data Lakes, and Data Lakehouses. In Data Warehousing, Data Lakes and Data Lakehouses. SAGE Publications, Inc. https://doi.org/10.4135/9781071937587