JNTM JOURNAL of NEW TECHNOLOGY and MATERIALS

Journal of New Technology and Materials

Vol. 16, No. 2 (2026), 31–40

RESEARCH ARTICLE

The Role of Databases in the Development of the Internet of Things (IoT): Data Storage and Processing Requirements

Mykola Madinov[1,*], Nina Petrukha[2], Serhii Petrukha[3], Halyna Melnyk[4] and Vasyl Melnyk[5]

[1]Department of Mobile and Video Information Technologies, State University of Information and Communication Technologies, Kyiv, Ukraine. [2]Department of Management in Construction, Kyiv National University of Construction and Architecture, Kyiv, Ukraine. [3]Department of Transport and Logistics, West Ukrainian National University, Ternopil, Ukraine. [4]Department of Applied Mathematics and Information Technologies, Faculty of Mathematics and Computer Science, Yuriy Fedkovych Chernivtsi National University, Chernivtsi, Ukraine. [5]Department of Mathematical Modelling, Faculty of Mathematics and Computer Science, Yuriy Fedkovych Chernivtsi National University, Chernivtsi, Ukraine

*Correspondence: nmadinov@gmail.com

Received date: April 28, 2026; Accepted date: September 20, 2026

Abstract

The article examines the role of database systems in the development of Internet of Things (IoT) ecosystems, with particular emphasis on the requirements for data storage, processing, and management amid the rapid increase in the number of connected devices and the volume of sensor-generated information. Traditional centralized relational databases are shown to be increasingly unable to satisfy the requirements of modern IoT systems because of high workloads, the need for horizontal scalability, real-time streaming data processing, and support for distributed architectures. The purpose of the study is to provide a comprehensive analysis of current trends in the development of databases and data-processing architectures in the IoT environment and to identify key technological approaches to the organization of IoT data storage and analytics. The research methodology combines bibliometric analysis with a qualitative review of the literature. The Scopus database served as the data source, from which 328 relevant publications covering 2012–2025 were selected using a specialized search query. VOSviewer software was used to visualize thematic relationships, keyword co-occurrence patterns, and research trends. The results indicate the dominance of research areas such as edge computing, distributed database systems, cloud and fog computing, stream processing, big data, and artificial intelligence. The analysis further shows that modern IoT systems increasingly rely on hybrid edge–fog–cloud architectures and specialized time-series databases to achieve low latency, scalability, and efficient processing of continuous data streams. The growing importance of AI-based analytics, cybersecurity mechanisms, and blockchain solutions in distributed IoT environments is also demonstrated. Overall, the results confirm that the convergence of time-series databases, edge computing, and artificial intelligence is becoming a fundamental technological basis for the next generation of IoT infrastructure.

Keywords: internet of things; time-series databases; edge computing; fog computing; cloud computing; distributed databases.

1 Introduction

The Internet of Things (IoT) is one of the key drivers of the Fourth Industrial Revolution and is transforming industry, healthcare, transportation, and the urban economy. According to analytical reports, the number of active IoT devices in 2025 exceeded 21 billion. Moreover, the annual volume of data generated by these devices already reaches tens of zettabytes and continues to grow rapidly [1, 2]. This rapid expansion in data volume creates fundamentally new requirements for information-storage, processing, and management systems. These requirements extend well beyond the capabilities for which traditional relational databases were originally designed.

© 2026 the Author(s). Published by Journal of New Technology and Materials. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0).

Despite the substantial body of research on IoT architectures, the evolution of database technologies in relation to the specific requirements of the Internet of Things has not been examined systematically enough. Most existing studies focus either on the technical aspects of edge and fog computing or on individual storage solutions, without comprehensively considering the relationships among database types, architectural paradigms, and the operational requirements of real-world IoT systems. In addition, relatively few studies combine bibliometric analysis with an in-depth technological review of modern time-series databases and hybrid edge–cloud solutions.

To address this gap, the purpose of this study is to comprehensively analyze the role of database systems in the development of the Internet of Things, identify the key requirements for storing and processing IoT data, and assess the evolution of research trends in this area.

The scientific novelty of the study lies in three interconnected aspects: the systematic combination of bibliometric analysis with a qualitative review of technological solutions; the generalization of functional and architectural requirements for IoT databases; and the identification of the temporal evolution of research priorities from infrastructure-oriented approaches toward intelligent, distributed, and security-oriented solutions.

2 Literature review

Over the past decade, the Internet of Things (IoT) has become one of the key technologies for the digital transformation of industry, transport, medicine, and urban infrastructure. The basic concepts of IoT were formed in the fundamental work of Atzori et al. [3], where the authors defined IoT as the integration of physical objects, sensor networks, and information and communication technologies into a single data exchange environment. The researchers emphasized that the main feature of IoT is the constant interaction between the physical and digital worlds, which leads to the continuous generation of large amounts of information. It was this characteristic that later became the main reason for the emergence of new approaches to data storage and analysis.

The further development of IoT systems has led to a sharp increase in both the number of connected devices and the volume of generated data. According to Cisco [4], global IP traffic associated with connected devices has expanded with the widespread adoption of smart devices, industrial automation, and data-intensive digital services. This growth places substantial pressure on storage systems and network infrastructure. Consequently, classical centralized models of information processing are increasingly inadequate for high-load IoT scenarios that require scalable ingestion, distributed processing, and rapid response.

The features of IoT data are analyzed in detail in the work of Hu et al. [5]. The authors note that most IoT data belongs to the category of time-series data, since each value is time-bound and generated in streaming mode. The study highlights that traditional relational databases are poorly adapted to handle high-frequency sensory data streams. Since they have the inherent complexity of scaling and significant resource costs for indexing. The authors also point out that IoT systems are characterized by high data-generation rates, uneven flows, heterogeneity of formats, as well as the need for real-time information processing.

The problems of the evolution of database architectures for IoT are actively considered in modern studies of distributed systems. Wang et al. [6] note that industrial IoT systems can include thousands of devices and millions of sensors that generate billions of records every day. The authors emphasize that traditional DBMS do not provide sufficient efficiency to work with such loads, especially in cases of delayed data arrival, schema evolution and highly concurrent ingestion. To solve these problems, the researchers proposed Apache IoTDB, a specialized time-series database that uses its own TsFile format and optimized data compression mechanisms. According to the results of experiments, the system demonstrated performance of up to 10 million inserts per second and performed complex aggregation queries in less than 100 ms.

The need to use specialized time-series databases is also confirmed in the work of de Oliveira et al. [7]. A systematic literature review of approaches to the compression of IoT time series was conducted here. The authors found that modern compression-oriented storage algorithms allow you to reduce the amount of data several times without significant loss of accuracy. This is especially important for edge devices, as they have limited memory and power consumption. The researchers emphasize that effective compression is one of the key factors in scaling IoT systems.

In parallel with the growth of IoT data, the transformation of computing architectures is taking place. One of the most influential concepts in recent years has been edge computing. In the foundational study by Shi et al. [8], edge computing is defined as an approach in which data processing is moved closer to the sources at which data are generated. The authors emphasize that this allows minimizing latency, optimizing the use of bandwidth and ensuring the operation of real-time systems. A further development of this concept is presented by Khan et al. [9], who consider edge computing a critical element in the operation of large-scale IoT systems. The researchers note that centralized cloud processing is becoming increasingly ineffective due to data transmission delays and excessive network load.

Similar conclusions were reached by Gkonis et al. [10], who investigated IoT–edge–cloud continuum systems. The authors note that modern IoT architectures are increasingly using hybrid approaches, where part of the calculations are performed at the edge level, and long-term storage and complex analytics are implemented in cloud data centers. The study notes that edge computing significantly increases fault tolerance. The technology makes it possible to process critical data even with an unstable network connection.

The integration of cloud, fog, and edge computing is also analyzed by Kuchuk and Malokhvii [11]. The authors note that fog computing acts as an intermediate layer between edge devices and the centralized cloud. This approach is able to provide a balance between the speed of local processing and the scalability of cloud systems. The researchers emphasize that multi-tiered architectures are becoming necessary for industrial IoT scenarios that have high performance requirements.

Andriulo et al. [12] in a review paper compare cloud, edge, and hybrid IoT architectures and show that centralized cloud systems have limitations in scenarios where real-time data processing is required. The authors emphasize that edge computing makes it possible to perform calculations closer to data sources, which reduces latency and increases the efficiency of system response. It is also noted that hybrid edge–cloud approaches are the most promising for IoT, as they combine the scalability of the cloud with the efficiency of local data processing.

Hasan and Idrees [13] emphasize that edge computing has actually become an integral part of modern AIoT systems. The authors note that the use of artificial intelligence in IoT requires local data processing, as the transfer of all flows to cloud centers creates significant bandwidth utilization and latency problems. The paper notes that the future development of IoT will be closely related to distributed AI processing and edge intelligence.

The development of distributed intelligence on edge-to-cloud continuum is explored in the work of Rosendo et al. [14]. The authors analyze modern approaches to distributed execution of AI models in heterogeneous environments. The researchers emphasize that edge AI allows not only to reduce latency, but also to increase the level of data privacy due to the local processing of sensitive information.

An important place in modern IoT systems is occupied by data streaming. Hmissi and Ouni [15] analyze MQTT as one of the basic IoT communication protocols. The authors note that MQTT provides low data overhead and works efficiently in low-bandwidth environments. At the same time, the classic broker-based MQTT architecture has limitations on horizontal scaling, so modern research is increasingly focused on distributed MQTT brokers.

The issues of fault tolerance and distributed stream processing are explored in Javed et al. [16], which presents the CEFIoT architecture based on Apache Kafka and Kubernetes. The authors demonstrate that the use of distributed streaming platforms allows for high availability of IoT services even in cases of failures of individual nodes. The study also highlights the importance of automated orchestration for edge–cloud systems.

The reliability of edge architectures is further analyzed by Diván et al. [17]. The authors conducted a systematic mapping study of contemporary approaches to real-time reliability monitoring in distributed environments. The researchers note that edge computing is characterized by high infrastructure dynamism and requires adaptive fault-tolerant mechanisms to ensure the stable operation of services.

A separate research area concerns the use of artificial intelligence to analyze IoT data streams. Trilles et al. [18] examine the concept of the Artificial Intelligence of Things (AIoT) and analyze anomaly-detection techniques for streaming data. The authors note that traditional statistical methods can become insufficient when multidimensional IoT data exhibit complex temporal dependencies, nonlinear relationships, and rapidly changing operating conditions.

Similar conclusions were reached by Chatterjee and Ahmed [19]. They conducted a survey of modern methods for detecting anomalies in IoT systems. The researchers emphasize that anomalies can be associated with both technical malfunctions of sensors and cyberattacks. The paper notes that deep learning algorithms have significantly higher accuracy compared to rule-based approaches in complex streaming environments.

Liu et al. [20] also focus on the use of unsupervised deep learning to analyze multidimensional time series. The authors note that modern IoT systems generate highly correlated sensor streams, which are already difficult to analyze with traditional algorithms. The study emphasizes that the effectiveness of AI analytics is closely related to the speed of access to historical time-series data.

Information security issues in edge-IoT environments are explored by Xue et al. [21]. The authors note that the distributed nature of edge computing increases the number of potential points of attack. It also complicates centralized access control. The paper analyzes modern access control models, as well as blockchain-based approaches for IoT data protection.

Questions concerning IoT data quality are addressed by Isa and Emran [22]. The authors emphasize that large IoT networks often generate noisy observations, duplicate records, and incomplete data, all of which can adversely affect the effectiveness of AI-based analytics. They consequently emphasize the need for automated data-validation frameworks and robust mechanisms for anomaly filtering and data-quality control.

Thus, the analysis of modern literature shows that the further development of the Internet of Things is inextricably linked with the evolution of distributed databases, edge computing technologies and AI-driven analytics. Current research unanimously confirms the benefits of specialized time-series databases for efficient IoT data processing, as well as the urgent need to use hybrid edge–cloud architectures to ensure scalability, low latency, and real-time processing. At the same time, a number of important issues remain insufficiently studied, including the standardization of IoT data management, increasing the fault tolerance of distributed systems, and optimizing AI algorithms in heterogeneous environments.

3 Materials and methods

The study adopts a combined methodological design that integrates systematic bibliometric analysis with a qualitative review of the scientific literature. The main goal was to identify current trends in the development of database systems and data processing architectures in the Internet of Things (IoT) ecosystem, with a special focus on requirements for storage, scalability, real-time processing, and integration with edge, fog, and cloud computing.

The source of bibliographic data for the study was the Scopus database [23]. The search was conducted in May 2025 using a structured query applied to the TITLE-ABS-KEY fields. The query included four thematic blocks, which were combined by logical operators.

Filters were applied to the results by document type (Article, Conference Paper, Conference Review), language (English), and publication period (2012–2025). The search yielded 328 relevant documents, which formed the dataset for the subsequent bibliometric and qualitative analyses.

The resulting dataset was analyzed in two stages. At the first stage, bibliometric analysis was performed using VOSviewer software (version 1.6.20). A keyword co-occurrence network and thematic clusters were constructed, an overlay visualization of the average publication year was generated, and the citation characteristics of key terms were examined. At the second stage, a qualitative content analysis of the selected publications was conducted to identify the principal technological approaches, architectural solutions, advantages, and limitations of existing IoT data-storage and processing systems.

Thus, the methodology used made it possible to combine the quantitative analysis of the scientific landscape with a qualitative synthesis of knowledge in the subject area under study.

4 Results

The purpose of this section is to analyze the scientific literature on data storage and processing architectures in the Internet of Things (IoT) ecosystem, with a special emphasis on the use of time-series databases, distributed storage systems, and streaming technologies in combination with edge, fog, and cloud computing.

Modern IoT systems are characterized by an exponential growth in the amount of data generated by sensor devices, as well as the need to process them in real time. This creates complex requirements for: scalability of storage systems; efficiency of stream processing; minimizing latency; edge–cloud architectures integration.

Accordingly, a research question arises: what technological approaches to storing and processing IoT data (in particular, database systems and stream processing) dominate the scientific literature, and how do they integrate with edge/fog/cloud architectures?

To solve the problem, a query in the Scopus database [23] was used, built on the principle of multidimensional filtering: TITLE-ABS-KEY (((("internet of things" OR IoT) AND ("time series database" OR TSDB OR "distributed database" OR "data storage" OR "database systems") AND ("data processing" OR "stream processing" OR "real-time processing" OR "data management") AND ("edge computing" OR "fog computing" OR "cloud computing"))) AND (LIMIT-TO (DOCTYPE , "cp") OR LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE , "cr"))

The query is structured in the form of four thematic blocks:

(1) IoT as a domain of research ("Internet of Things" OR IoT). This block ensures that all selected documents belong to the IoT domain and excludes irrelevant work from related areas (e.g. pure cloud computing or general distributed systems).

(2) Data Storage Layer ("time series database" OR TSDB OR "distributed database" OR "data storage" OR "database systems"). This block provides a focus on: time-series databases (IoTDB, InfluxDB, TimescaleDB); distributed storage systems; IoT data management architecture. It is this level that corresponds to the key topic of the study – requirements for storing IoT data.

(3) Processing Layer ("data processing" OR "stream processing" OR "real-time processing" OR "data management"). This component makes it possible to include research related to: streaming data processing (Kafka, Flink, Spark); real-time analytics; IoT data flow management systems. This block is necessary for analyzing real-time data processing in IoT environments.

(4) edge–cloud Continuum ("edge computing" OR "fog computing" OR "cloud computing"). This block provides the inclusion of works that describe: edge computing as local data processing; fog computing as an intermediate level; cloud computing as a centralized infrastructure. This makes it possible to explore the architectural evolution of IoT systems.

(5) Document type filtering, in particular by the LIMIT-TO (DOCTYPE, "cp") parameter. Restrictions on conference papers (CP) are used for the purpose of: capturing the most relevant and quickly updated research in the IoT/edge/DB field; taking into account the fact that in this field a significant part of innovations is first published at conferences (IEEE, ACM, etc.); ensuring the representativeness of modern architectural solutions.

The resulting set of 328 documents is appropriate from the point of view of bibliometric analysis, since: the sample is large enough for statistical processing (trend analysis, clustering); but limited enough to provide thematic relevance; covers the period of active development of edge–IoT–TSDB systems (2012–2025), consistent with the defined search interval.

This query makes it possible to form a single research plane: IoT data lifecycle: generation → storage → processing → architecture.

Thus, it provides a comprehensive analysis of: databases (TSDB, distributed DB); stream processing; edge/cloud/fog architectures; real-world IoT workloads.

The Scopus query used in this study makes it possible to obtain a representative sample of studies reflecting the current state of development of IoT systems in the context of data storage and processing. This provides an opportunity for further bibliometric and content analysis to identify the main technological trends and architectural approaches.

As part of the study, the VOSviewer [24] software was used to build a keyword matching map based on a sample of publications from the Scopus database. To visualize the structure of the research area, a minimum threshold of 100 keyword occurrences was selected. The resulting network made it possible to identify six main thematic clusters (Figure 1).

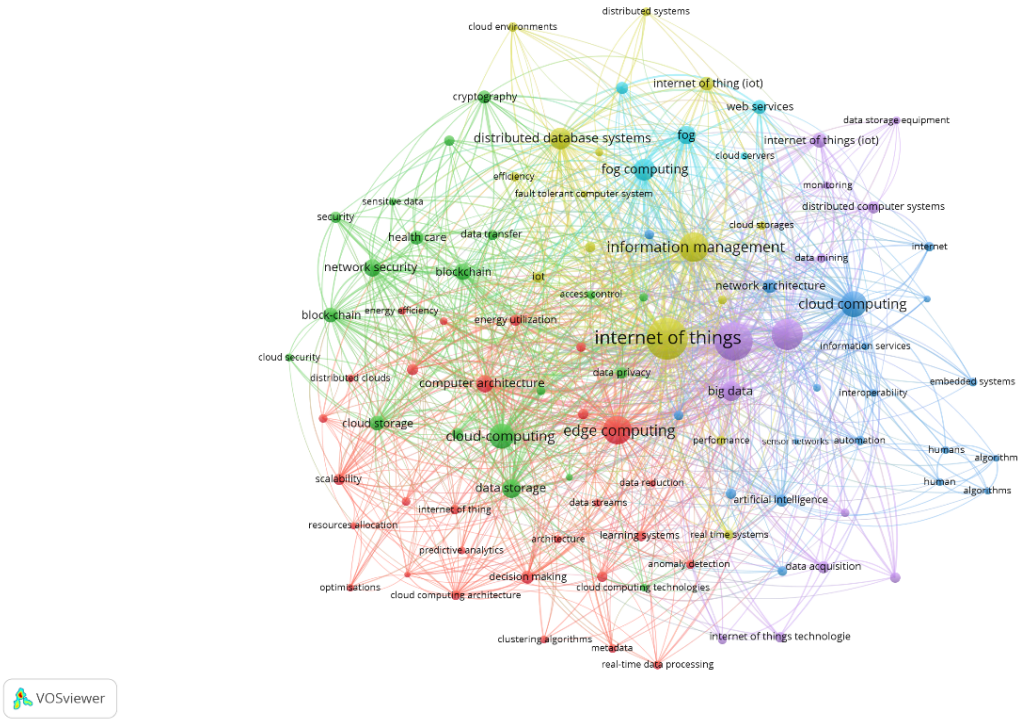


Figure 1. Keyword co-occurrence network and cluster structure of IoT data management research (minimum occurrences = 100)
Source: Built by the authors using the VOSviewer app [24]

Cluster 1: Edge computing and real-time data processing. The cluster brings together research that is related to edge computing as well as real-time data processing. It includes keywords such as edge computing, real-time data processing, predictive analytics, anomaly detection, resource allocation, and scalability. This direction reflects the current trend, which is to move computing processes closer to data sources. This makes it possible to minimize delays and increase the efficiency of IoT systems. Much attention is paid here to resource optimization and energy efficiency. These are critical factors for scalable distributed environments.

Cluster 2: Security, privacy, and blockchain. The cluster focuses on security and data protection issues in IoT. It includes terms such as blockchain, security, data privacy, access control, and cryptography. This direction emphasizes the need to ensure reliable storage and transmission of data in distributed networks. The use of blockchain technologies indicates a growing interest in decentralized trust mechanisms, as well as data integrity and transaction transparency in IoT environments.

Cluster 3: Artificial intelligence and intelligent networks. The cluster covers research related to the integration of artificial intelligence into IoT systems. Among the keywords are artificial intelligence, machine learning, algorithms, sensor networks, and network architecture. This direction reflects the transition to the increasing intelligence of IoT systems. Here, systems are able to independently analyze data, as well as make decisions and adapt to changes in the environment. The combination of machine learning algorithms with network technologies opens up new opportunities for automation and improved data processing efficiency.

Cluster 4: Basic IoT systems and distributed databases. The cluster forms the basis of the entire subject

area. It includes concepts such as the Internet of Things, distributed database systems, information management, and fault tolerance. It reflects the fundamental principles of building IoT systems. In particular, it is the use of distributed databases to ensure reliability, scalability, and continuity of information processing. It is this cluster that is the core around which other thematic areas are formed.

Cluster 5: Data management and big-data analytics. The cluster is dedicated to data management processes in the IoT environment. It includes big data, data handling, data acquisition, data mining, and digital storage. This area covers the full life cycle of data – from data collection to analysis and visualization. It is directly related to the topic of databases. Since the efficient storage and processing of large amounts of information is a key condition for the functioning of modern IoT solutions.

Cluster 6: Cloud and fog computing. The cluster reflects the IoT infrastructure layer. It includes concepts such as cloud computing, fog computing, cloud services, and web services. It demonstrates the role of cloud platforms in providing storage, scalability, and computing power for IoT systems. At the same time, the concept of fog computing acts as an intermediate layer between the cloud and the edge, which makes it possible to optimize data processing and reduce latency.

Thus, the conducted cluster analysis shows that modern research in the field of the Internet of Things is structured around three key dimensions: infrastructure, data processing and their mining. The results obtained confirm the decisive role of storage systems and databases in the development of effective and scalable IoT solutions.

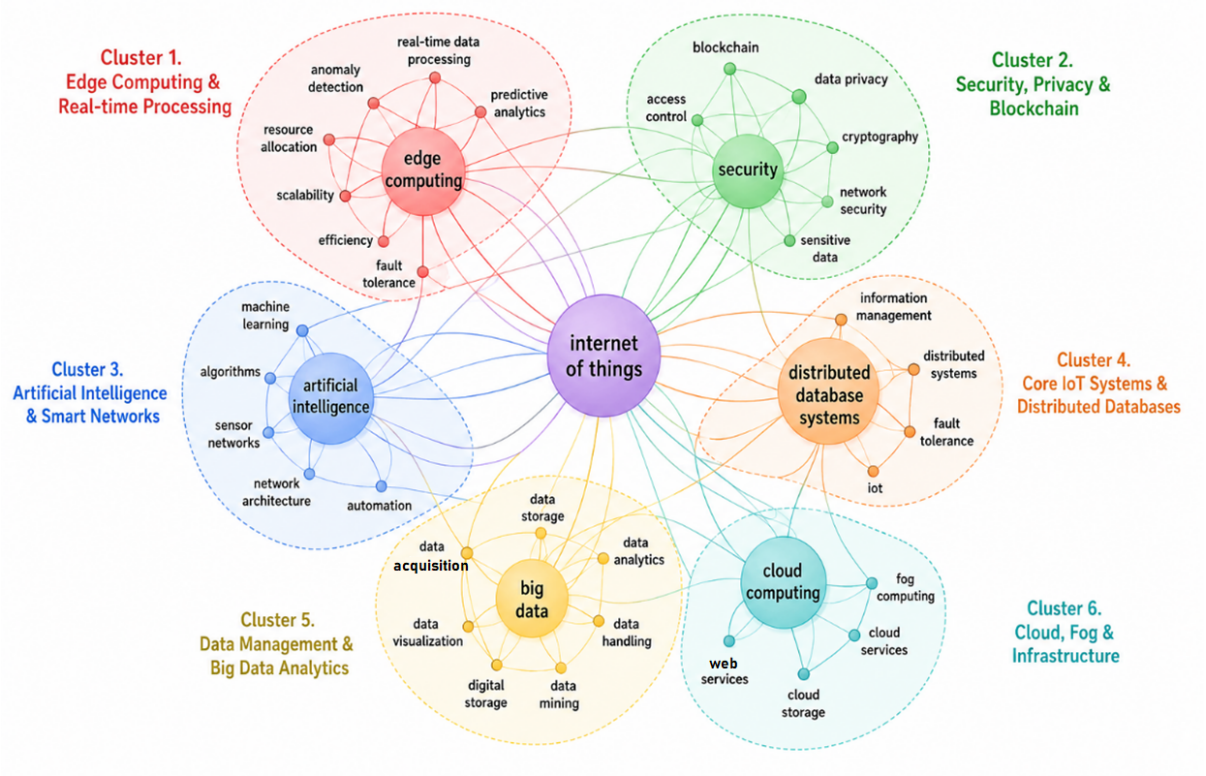


Figure 2. Co-occurrence network and cluster structure of IoT research domains: storage, processing, and edge–cloud integration
Source: built by the authors

Figure 2 summarizes all clusters of IoT research in the form of a single network of interconnections between key topics. The central element is the "Internet of Things", from which the main thematic areas radially diverge – edge computing, security, AI, distributed databases, big data and cloud/fog computing.

Graphically, it can be seen that clusters form three large levels: infrastructure (cloud–edge–fog), computational and analytical (AI, big data, stream processing) and system-data (distributed databases, storage, data management). The tightest links are observed between edge computing, security, and cloud computing, reflecting their central role in contemporary IoT architectures.

Each cluster has a distinct thematic core, but at the same time, all groups are closely interconnected, which emphasizes the integrated nature of the development of IoT as a single ecosystem "data–processing–infrastructure".

A bibliometric analysis of 328 publications indexed in the Scopus database dedicated to data management in the Internet of Things (IoT) revealed a dense and well-structured network of keyword co-occurrences built using VOSviewer software [24]. The resulting dataset covers a large number of interrelated terms. They reflect key areas of research in the field of storage, distributed systems, edge–cloud architectures and intelligent analytics.

The most dominant keywords are "internet of things" (199 appearances, total strength of connections – 1482), "digital storage" (169; 1183), "edge computing" (92; 661), "cloud computing" (80; 570) and "big data" (45; 384). These nodes form the central core of the network. They reflect the underlying technological infrastructure of the IoT ecosystem.

A structural analysis of the network shows a pronounced centralization around the concepts of data management and infrastructure (in particular IoT systems, distributed databases, and cloud/edge computing). High values of link strength for the terms data handling (757), information management (708) and cloud computing (570) indicate that data lifecycle management is one of the key areas of modern research in this area.

To assess the dynamics of research development over time, temporal visualization (overlay visualization) was built with a minimum keyword frequency of 100 (Figure 3). The color scale displays the average year of publications for each term.

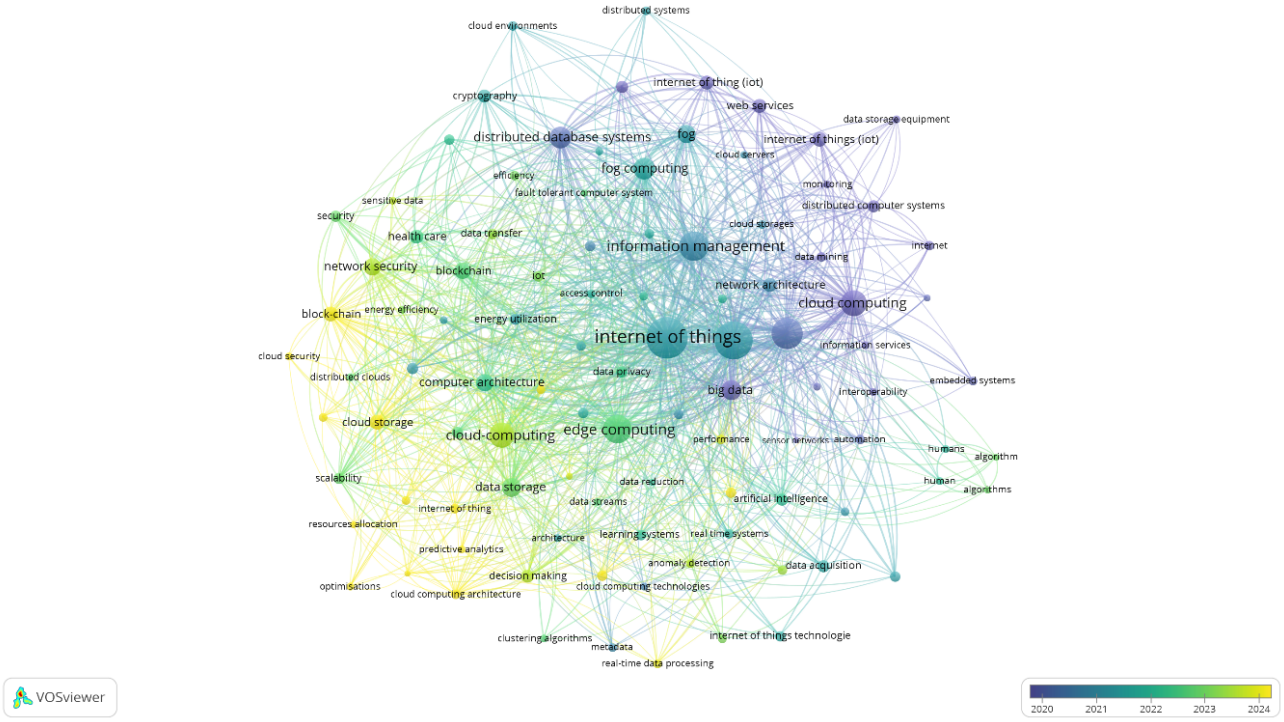


Figure 3. Overlay visualization of the keyword co-occurrence network showing the temporal evolution of research topics in IoT data management (minimum occurrences = 100)

Source: Built by the authors using the VOSviewer app [24]

The analysis of Figure 3 indicates a clear evolution of scientific interests. Earlier studies (2020–2021) focused mainly on basic concepts. These are constructions such as: cloud computing, distributed database systems, data storage and Internet of Things. Since 2022, there has already been a significant increase in attention to edge computing, fog computing, real-time data processing, and security. The most recent topics (2023–2024) are related to blockchain, machine learning, anomaly detection, data privacy, as well as the integration of artificial intelligence into IoT systems. Such dynamics indicate a shift from general infrastructure solutions to intelligent, security-oriented and real-time data processing systems.

Citation analysis additionally emphasizes the scientific impact of certain areas. Terms such as big data (29.24 average citations), data mining (82.18), and security (77.53) demonstrate the highest level of academic impact. This confirms the importance of research in the field of data analytics and cybersecurity for IoT systems.

In general, the obtained bibliometric indicators confirm that modern research on the Internet of Things is characterized by the transition from classical distributed architectures to edge-oriented, AI-oriented and real-time systems. The key factors here are scalability, minimizing latency, and data security.

5 Discussion

The results of the study provide a comprehensive picture of the evolving role of database systems in the development of Internet of Things (IoT) ecosystems. Based on a systematic literature review and bibliometric analysis of 328 publications indexed in the Scopus database, it can be seen that data storage and processing architectures have become a central element in solving the key challenges of IoT environments, namely high data-generation rates, heterogeneity, real-time

requirements, and scalability.

The key requirements for data storage and processing systems in IoT systems are summarized in Table 1.

Table 1. Key requirements for data storage and processing in IoT systems

Requirement	Description	Recommended solution	Main sources
High write speed	Generation of thousands to millions of records per second	Time-series databases (IoTDB, InfluxDB)	Wang et al. [6]
Low latency	Real-time data processing (< 100 ms)	Edge computing + Fog layer	Shi et al. [8]; Gkonis et al. [10]
Scalability	Support for thousands of devices and millions of sensors	Distributed databases + Cloud scaling	Wang et al. [6]
Energy efficiency	Limited resources of edge devices	Data compression + Edge preprocessing	De Oliveira et al. [7]
Security and privacy	Protection of sensitive data from unauthorized access	Blockchain, encryption, access control	Xue et al. [21]
Fault tolerance	System operation in case of failure of individual nodes	Fault-tolerant fog/edge architectures	Javed et al. [16]
Data heterogeneity	Support for various formats and protocols	Middleware + Schema evolution support	Hu et al. [5]

Source: built by the authors

5.1 Data architectures and storage systems

As noted by Shi et al. [8], Khan et al. [9], Gkonis et al. [10], and Andriulo et al. [12], centralized data processing in the cloud has significant limitations in the context of latency in IoT applications with stringent real-time requirements. Bibliometric analysis confirms this trend. The term "edge computing" is one of the most central and closely related keywords. This transition demonstrates the practical need to move computing closer to data sources. This is done to minimize latency, optimize network bandwidth usage, and ensure that systems are running in real-time. Fog computing accordingly strengthens this continuum, as it is an intermediate level. It strikes a balance between the speed of on-premises processing and the scalability of cloud resources [11].

The study confirms that traditional relational databases are increasingly unable to satisfy to the requirements of IoT workloads. Time-series databases (TSDBs), particularly Apache IoTDB, are gaining considerable popularity due to their ability to efficiently operate at high write speeds, schema evolution, and complex temporal queries [6]. The bibliometric map demonstrates the tight connections between the concepts of "distributed database systems", "data storage" and "big data", which indicates an active search for specialized solutions for storing IoT data, which are characterized by high frequency, time reference and often unevenness or noise. Data compression techniques, as well as edge-level preprocessing [7], are also gaining importance. Since they are important strategies for reducing the volume of data storage and transmission.

5.2 Security and intelligent analytics

Security and data protection form a distinct and prominent research cluster that is closely connected with blockchain, access control, and cryptography. This pattern is consistent with the findings of Xue et al. [21] and Hasan and Idrees [13]. These studies emphasize that the distributed nature of edge computing can substantially expand the attack surface, particularly in sensitive application domains such as healthcare and industrial IoT. The high citation rates associated with security-related keywords further emphasize the academic and practical significance of this research direction.

The convergence of IoT with artificial intelligence (AIoT) can be clearly traced. The key terms "machine learning", "anomaly detection" and "predictive analytics" have close links to edge computing and real-time data processing. This reflects the shift towards intelligent, self-learning systems capable of local decision-making [18–20]. Performing anomaly detection and predictive maintenance directly at the edge level significantly increases the autonomy of systems. It also reduces reliance on constant communication with the cloud.

5.3 Implications, limitations, and future research

Theoretically, the study contributes to an understanding of IoT as a sociotechnical ecosystem in which data management is not merely an auxiliary function but a defining capability that affects system performance, reliability, scalability, and security. The findings point to three closely connected architectural dimensions: infrastructure represented by edge–fog–cloud resources, data management represented by time-series and distributed database systems, and intelligence represented by AI and machine-learning methods.

Although the study yields useful results, several limitations should be acknowledged. The analysis is restricted to publications indexed in the Scopus database, including articles and conference-related documents, and may therefore exclude relevant studies indexed elsewhere. In addition, although bibliometric analysis is effective for identifying broad trends, thematic structures, and relationships among research topics, it does not by itself reveal the full technical depth, implementation details, or comparative performance of individual solutions.

Promising areas of further research are:

– empirical comparison of performance of different TSDBs in real-world industrial IoT scenarios;
– development of standardized frameworks for ensuring the quality of IoT data;

– creation of energy-efficient AI models for edge devices with limited resources;
– study of interoperability between different edge–fog–cloud platforms.

Therefore, database systems in IoT ecosystems are no longer passive storage systems. They have evolved into active, intelligent, and distributed components. They define the capability, scalability, and trust of modern IoT applications. Further convergence of time-series databases, edge computing, and artificial intelligence can form the foundation for the next generation of IoT infrastructure.

6 Conclusions

The study shows that, in modern IoT ecosystems, databases no longer function merely as passive information repositories. They have evolved into active, intelligent, and distributed components that substantially influence the performance, scalability, reliability, and level of trust that can be achieved across IoT systems.

The results confirm an important evolution in approaches to data management, from traditional centralized cloud solutions toward hybrid edge–fog–cloud architectures. Specialized time-series databases, combined with edge-computing technologies and artificial intelligence, are increasingly becoming key elements for efficiently storing, managing, and processing large volumes of sensor data in real time.

Thus, the further development of the Internet of Things will largely depend on the creation of smart, adaptive, and secure data storage and processing systems. The convergence of time-series databases, edge computing, and AI-based analytics is therefore forming the foundation of the next generation of IoT infrastructure, which must satisfy increasingly demanding requirements for processing speed, reliability, scalability, and information protection as data volumes continue to grow.

Author contributions

All authors contributed equally and reviewed and approved the final version of the manuscript and agreed to be accountable for the integrity of the work.

Conflict of interest

The authors declare that they have no conflicts of interest related to this research.

Data availability statement

The bibliographic data analyzed in this study were obtained from the Scopus database using the search strategy described in the Materials and methods section. The data supporting the reported bibliometric findings are described within the manuscript, and additional analytical files may be obtained from the corresponding author upon reasonable request, subject to applicable database licensing restrictions.

Funding

This research received no external funding.

References

[1] Sinha, S. (2025, October 28). State of IoT 2025: Number of connected IoT devices growing 14% to 21.1 billion globally. *IoT Analytics*. https://iot-analytics.com/number-connected-iot-devices/

[2] Ericsson. (2025). *Ericsson mobility report: November 2025*. https://www.ericsson.com/en/reports-and-papers/mobility-report/reports

[3] Atzori, L., Iera, A., & Morabito, G. (2010). The Internet of Things: A survey. *Computer Networks, 54*(15), 2787–2805.

[4] Cisco Systems. (2020). *Cisco annual internet report (2018–2023) white paper*. https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html

[5] Hu, C., Sun, Z., Li, C., Zhang, Y., & Xing, C. (2023). Survey of time series data generation in IoT. *Sensors, 23*(15), Article 6976.

[6] Wang, C., Qiao, J., Huang, X., Song, S., Hou, H., Jiang, T., Rui, L., Wang, J., & Sun, J. (2025). Apache IoTDB: A time series database for large scale IoT applications. *ACM Transactions on Database Systems, 50*(2), Article 7, 1–45.

[7] de Oliveira, M. A., da Rocha, A. M., Puntel, F. E., & Cavalheiro, G. G. H. (2023). Time series compression for IoT: A systematic literature review. *Wireless Communications and Mobile Computing, 2023*(1), Article 5025255.

[8] Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. *IEEE Internet of Things Journal, 3*(5), 637–646.

[9] Khan, W. Z., Ahmed, E., Hakak, S., Yaqoob, I., & Ahmed, A. (2019). Edge computing: A survey. *Future Generation Computer Systems, 97*, 219–235.

[10] Gkonis, P., Giannopoulos, A., Trakadas, P., Masip-Bruin, X., & D'Andria, F. (2023). A survey on IoT-edge–cloud continuum systems: Status, challenges, use cases, and open issues. *Future Internet, 15*(12), Article 383.

[11] Kuchuk, H., & Malokhvii, E. (2024). Integration of IoT with cloud, fog, and edge computing: A review. *Advanced Information Systems, 8*(2), 65–78.

[12] Andriulo, F. C., Fiore, M., Mongiello, M., Traversa, E., & Zizzo, V. (2024). Edge computing and cloud computing for Internet of Things: A review. *Informatics, 11*(4), Article 71.

[13] Hasan, B. T., & Idrees, A. K. (2024). Edge computing for IoT. *arXiv* [Preprint], arXiv:2402.13056. https://arxiv.org/abs/2402.13056

[14] Rosendo, D., Costan, A., Valduriez, P., & Antoniu, G. (2022). Distributed intelligence on the edge-to-cloud continuum: A systematic literature review. *Journal of Parallel and Distributed Computing, 166*, 71–94.

[15] Hmissi, F., & Ouni, S. (2022). TD-MQTT: Transparent distributed MQTT brokers for horizontal IoT applications. In *2022 IEEE 9th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT)* (pp. 479–486). IEEE.

[16] Javed, A., Heljanko, K., Buda, A., & Främling, K. (2018). CEFIoT: A fault-tolerant IoT architecture for edge and cloud. In *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)* (pp. 813–818). IEEE.

[17] Diván, M. J., Shchemelinin, D. A., Carranza, M. E., Martinez-Spessot, C. I., & Buinevich, M. V. (2023). Real-time reliability monitoring on edge computing: A systematic mapping. *Informatics and Automation, 22*(6), 1243–1295.

[18] Trilles, S., Hammad, S. S., & Iskandaryan, D. (2024). Anomaly detection based on Artificial Intelligence of Things: A systematic literature mapping. *Internet of Things, 25*, Article 101063.

[19] Chatterjee, A., & Ahmed, B. S. (2022). IoT anomaly detection methods and applications: A survey. *Internet of Things, 19*, Article 100568.

[20] Liu, Y., Zhou, Y., Yang, K., & Wang, X. (2023). Unsupervised deep learning for IoT time series. *IEEE Internet of Things Journal, 10*(16), 14285–14306.

[21] Xue, T., Zhang, Y., Wang, Y., Wang, W., Li, S., & Zhang, H. (2024). Edge computing for IoT: Novel insights from a comparative analysis of access control models. *arXiv* [Preprint], arXiv:2405.07685. https://arxiv.org/abs/2405.07685

[22] Isa, S. N. B. M., & Emran, N. A. (2024). Data quality in IoT open datasets: A methodological review. *Journal of Advanced Computing Technology and Application, 6*(1), 25–36.

[23] Elsevier. (n.d.). *Scopus* [Database]. https://www.scopus.com

[24] van Eck, N. J., & Waltman, L. (2023). *VOSviewer* (Version 1.6.20) [Computer software]. Centre for Science and Technology Studies, Leiden University. https://www.vosviewer.com