Serverless & Disaggregated Database Architectures: A Systematic review

Khagendra Mishra[1]; Suresh Gautam[2]

[1]Nepal Open University, Lalitpur, Nepal, MPhil in ICT Scholars.
[2]Nepal Open University, Lalitpur, Nepal, MPhil in ICT Faculty Member.

Corresponding Author: Khagendra Mishra*

Publication Date: 2026/07/29

Abstract: Review synthesizes research on "Serverless and disaggregated database architectures: performance optimization, cost efficiency, scalability, historical evolution, current trends, practical applications in industries, comparison with traditional database systems" to address the knowledge gap regarding how emerging paradigms reshape database management in cloud-native environments. The review aimed to taxonomies architectural designs, evaluate performance and cost-efficiency strategies, benchmark scalability, identify industry applications, and compare historical evolution with current trends. A systematic analysis of empirical studies, prototypes, and theoretical works published up to mid-2024 was conducted, focusing on cloud-native deployments leveraging technologies such as RDMA, persistent memory, and function orchestration. Key findings reveal that these architectures enable elastic scaling and significant cost reductions through pay-as-you-go models and resource pooling, while performance optimization benefits from AI-driven scheduling and hardware co-design; however, challenges persist in cold-start latency, orchestration complexity, and consistency management. Industry adoption spans finance, retail, and IoT, demonstrating operational gains but constrained by migration complexity and tooling maturity. The evolution from monolithic to serverless and disaggregated systems is marked by innovations in decoupled resource management and multi-cloud strategies. These findings collectively underscore the transformative potential and practical limitations of serverless and disaggregated databases, informing future research and guiding effective industrial adoption.

Keywords: *Serverless Architectures, Disaggregated Databases, Performance Optimization, Scalability, Cost Efficiency.*

How to Cite: Khagendra Mishra; Suresh Gautam (2026) Serverless & Disaggregated Database Architectures: A Systematic review. *International Journal of Innovative Science and Research Technology*, 11(7), 1792-1796. https://doi.org/10.38124/ijisrt/26jul1367

I. INTRODUCTION

Serverless architecture is a cloud model where applications are built using small functions that execute only when triggered, automatically scale, and require no server management. Developers only write code; the cloud provider handles everything else. Billing is based solely on actual usage[1]. Stream Cloud, a highly scalable and elastic stream processing engine that parallelizes continuous queries across shared-nothing clusters while preserving semantic correctness. It proposes a novel Operator set parallelization strategy and efficient load-balancing mechanisms, achieving significant scalability and low-overhead elasticity for large real-time data streams[2].

Alibaba's cloud-native database ecosystem including POLARDB, POLARDB-X, AnalyticDB, and the SDDP platform designed through hardware–software co-design, shared-storage and shared-nothing architectures, and autonomous management to achieve high performance, extreme scalability, and reliability. It highlights key technologies such as PolarFS, X-Engine, Parallel Raft, and ML-driven optimization that enable Alibaba to support massive real-world workloads like Singles' Day while significantly improving efficiency and reducing operational cost[3].

There are two main objectives of the Research.

- Objective 1: To analyze the limitations of traditional cloud database systems in terms of performance and scalability, focusing on how serverless and disaggregated architectures address these issues.
- Objective 2: To evaluate the effectiveness of serverless and disaggregated architectures in achieving performance optimization and cost efficiency for modern data-intensive applications.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul1367

➢ *Problem of Statement*

Despite the rapid evolution of cloud native data management systems, the practical adoption of serverless and disaggregated database architectures remains hindered by unresolved challenges in performance consistency, orchestration efficiency, and cost predictability. Current systems offer theoretical benefits such as elastic scaling, pay-as-you-go cost models, and decoupled resource management; however, empirical studies reveal persistent issues, including cold-start latency, complex coordination among disaggregated components, and difficulty maintaining transactional consistency under dynamic workloads. Moreover, industry deployments lack mature tooling and standardized benchmarks, preventing a clear comparison with traditional monolithic and cloud-hosted databases. These limitations create a critical gap in understanding how serverless and disaggregated designs can be optimized to support real world, large-scale, latency-sensitive applications.

Therefore, this research aims to systematically investigate the architectural, algorithmic, and operational factors that influence performance, consistency, and cost efficiency in serverless and disaggregated database systems. By developing new optimization models, orchestration strategies, and benchmarking frameworks, the study seeks to address foundational gaps that currently prevent these emerging paradigms from achieving reliable and predictable behavior in cloud-native environments. So research questions of this study will be:

- RQ1: How do serverless and disaggregated database architectures overcome the performance and scalability challenges faced by traditional monolithic cloud databases?
- RQ2: To what extent do serverless and disaggregated architectures improve performance optimization and cost efficiency in real-world data-processing environments?

II. RESEARCH METHODOLOGY

In order to ascertain the pertinent articles for this review, a comprehensive search encompassing titles, abstracts, and keywords was conducted within the Scopus database, Google scholar and Research Gate utilizing the search phrase "Serverless & Disaggregated Database Architectures." This search yielded a total of 251 results. These results encompass a variety of paper. Of these, our focus is specifically directed towards the peer-reviewed publications Serverless architectures, Disaggregated databases Performance optimization, Scalability, Cost efficiency Consequently, all other articles were excluded, resulting in a remaining total of 51. This database served as the foundational resource for subsequent review and analysis. Upon meticulous examination of the titles, abstracts, and keywords, accompanied by a focus on significant application domains, we ultimately identified 20 articles for the final revision. We have critically reviewed these 20 papers which have been published in prestigious international journals from the years 2010 to 2025. This paper uses a qualitative and comparative research method. It studies information from academic articles, cloud service providers, and benchmarking reports. It compares the architecture of serverless systems with disaggregated database systems. The evaluation is based on important factors such as latency, throughput, elasticity, storage efficiency, and operational cost in the field of Serverless & Disaggregated Database Architectures.

III. LITERATURE STUDY/REVIEW

Serverless computing is a promising paradigm that simplifies cloud development and reduces operational complexity. However, it still faces significant challenge especially related to tooling, security, performance, and portability[1]. Alibaba designs and deploys cloud-native database systems through shared-storage architectures, hardware–software co-design, and autonomous management to achieve high performance, massive scalability, and cost-efficient reliability for large-scale real-world workloads[3].A scalable framework that integrates IoT networks with block chain using smart contracts and off-chain storage to enhance security, privacy, authorization, and data integrity in resource-constrained healthcare systems, showing that the model improves trust, scalability, and real-time data management for medical applications[4]. BIoMT, a block chain-based Hyper ledger Fabric consortium architecture that secures IoMT healthcare data using smart contracts, NuCypher re-encryption, and serverless P2P communication, aiming to overcome privacy, integrity, and scalability limitations of traditional e-health systems. Experimental results show that BIoMT significantly reduces computational cost and enhances node performance, providing a more efficient, secure, and reliable infrastructure for medical data management[5].Disaggregated data center architectures, where resources like memory are accessed remotely, significantly degrade the query execution performance of production DBMSs like MonetDB and PostgreSQL, though some configurations show surprising minor degradation or even improvements by enabling better scaling[6].

Serverless lake house architectures, highlighting their capacity to deliver enterprise value by combining object storage flexibility with data warehouse performance through innovations like decoupled storage-compute paradigms, sophisticated metadata management for ACID guarantees, and elastic query processing, while also addressing challenges such as cold-start latency and complex joins, and exploring the transformative role of artificial intelligence in enhancing query optimization, workload management, and data governance, ultimately providing a comprehensive view of their impact on data management and identifying areas for continued innovation[7]. The implications of disaggregated data centers (DDCs) for relational database management systems (RDBMSs), highlighting that while DDCs offer operational benefits like independent resource upgrades and efficient utilization, a naive adaptation of traditional RDBMSs to these architectures can lead to significant performance degradation due to increased network communication and remote memory access, thus necessitating a rethinking of RDBMS design and proposing new hardware and OS primitives for improved performance and reliability[8].Service Level Agreements (SLAs) and prices for serverless query processing, proposing PixelsDB as an open-source prototype that uses dedicated architectural designs to support multiple service levels and achieve significant cost reductions while maintaining performance guarantees[9]. PolarDB Serverless is a novel

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul1367

cloud-native database that employs a disaggregated architecture, decoupling CPU, memory, and storage resources to provide on-demand provisioning, improved reliability, and faster failure recovery compared to traditional monolithic systems[10].The performance trade-offs between serverless (AWS Aurora Serverless v2) and provisioned (AWS Aurora Postgre SQL) database architectures under a benchmark workload, revealing that while serverless databases offer scalability and cost benefits, they introduce latency overhead due to cold starts and auto scaling mechanisms[11].

Efficient execution of data-intensive analytics workloads in serverless environments through direct data inter-function communication and optimized performance-cost efficient resource allocation policies[12].DSCS-Serverless, a serverless execution model that integrates a programmable domain-specific accelerator within computational storage to overcome data movement overheads in disaggregated datacenters and significantly improve performance, energy efficiency, and cost for serverless functions[13].The performance and cost characteristics of serverless cloud infrastructure, particularly AWS Lambda, S3, DynamoDB, and EFS, for large-scale data processing, identifying performance variability in networking and storage, and determining cost break-even points for efficient usage[14].

ArcaDB is a disaggregated query engine that leverages container technology to optimize performance and scalability in heterogeneous computational environments by placing operators at compute nodes that match their performance profiles, thus enabling efficient use of accelerators, improving query response time by up to 3.5x compared to shared-nothing symmetric arrangements, and keeping costs under control, particularly in cloud settings where resources are dynamically allocated[15]. Vexless is a pioneering vector database system designed for serverless architectures, leveraging cloud functions like AWS Lambda, Azure Functions, and Google Cloud Functions to provide elastic, serverless, and low-cost cloud computing for bursty and sparse workloads, It addresses the unique challenges of serverless environments, such as resource limitations, communication overhead, and cold-start times, through various performance optimizations[16].Starling is a query execution engine built on cloud function services, designed to provide interactive query latency at a lower total cost than traditional provisioned systems for ad-hoc analytical workloads, particularly on data stored in cloud object storage[17].Distributed shared-memory databases (DSM-DB) leveraging RDMA-enabled memory disaggregation (MD) as the future of database system innovation, outlining the challenges and opportunities for realizing high-performance OLTP main-memory databases in this architecture[18].

IV. ANALYSIS OF THE WORK

The reviewed literature on serverless and disaggregated database architectures reveals significant advancements in scalability, cost efficiency, and performance optimization, highlighting the transformative potential of these paradigms in cloud-native environments.

- *Performance Optimization:*

Some studies found that serverless and disaggregated architectures improve query latency and throughput through elastic scaling, AI-enhanced optimization, and fine-grained orchestration techniques[7][17][12]. Some papers highlighted challenges such as cold-start latency, communication overhead, and straggler mitigation that affect performance in serverless environments[7][11][16]. Several studies demonstrated that hardware co-design, RDMA, and persistent memory integration significantly enhance performance in disaggregated databases[10]. Performance models and scheduling algorithms have been developed to predict and optimize workload execution in serverless contexts[12][9].

- *Cost Efficiency:*

Some studies reported substantial cost reductions through pay-as-you-go models, resource pooling, and optimized resource allocation in serverless and disaggregated systems[16][14]. Hybrid architectures combining serverless functions with traditional VMs or containers achieve better cost-performance trade-offs for sustained workloads[15].Domain-specific accelerators embedded in storage devices reduce energy consumption and improve cost efficiency in serverless computing[13]. Cost savings are often balanced against overheads from cold starts and auto scaling, requiring careful workload characterization[14][11].

- *Scalability and Elasticity:*

Some studies emphasized independent and elastic scaling of compute, memory, and storage as a core advantage of disaggregated and serverless architectures[10][19]. Multi-primary and shared-storage designs enable high write throughput and read scale-out, enhancing scalability[19]. Fine-grained function orchestration and multi-agent reinforcement learning improve elasticity and resource utilization in geo-distributed and wide-area settings[20]. Challenges remain in managing coordination overhead and consistency in highly elastic environments[8].

Serverless databases excel in unpredictable, event-driven workloads due to automatic scaling and operational simplicity. Disaggregated architectures deliver superior performance for analytics by enabling multiple compute clusters on shared data. Both architectures significantly improve cost efficiency by eliminating over-provisioning. Scalability is enhanced across both models, but serverless provides better real-time elasticity, while disaggregated supports large-scale analytical workloads.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul1367

V. CONCLUSION

Recent research on serverless and disaggregated database architectures shows a major shift in how modern cloud databases are designed. These systems separate compute, memory, and storage so each can scale independently, making them highly elastic, cost-efficient, and scalable for cloud-native workloads.

By combining disaggregation with serverless computing pay as you go model, organizations can handle bursty or unpredictable workloads efficiently while reducing idle resource costs. Advanced networking tools like RDMA and persistent memory help overcome performance bottlenecks that traditionally slowed distributed databases.

Performance improves further through AI-driven query optimization, fine-grained function execution, and better coordination between hardware and software. However, challenges remain, such as cold-start delays, communication overhead between separated components, and difficulties in managing distributed transactions. Benefits also vary by workload type analytical queries often gain more than highly transactional workloads.

In terms of cost, serverless systems reduce waste but sometimes come with higher per-unit pricing and overhead from frequent function creation. Hybrid systems that blend serverless with traditional VMs can balance cost and performance.

Industries like finance, retail, IoT, and large cloud providers are already adopting these architectures, though migration complexity, data consistency issues, and immature tooling slow wider adoption. Overall, the trend clearly moves toward more AI-driven, multi-cloud, and edge-integrated database systems that are expected to shape the future of data management.

In summary, serverless and disaggregated database architectures offer compelling advantages in elasticity, cost, and scalability over traditional monolithic systems. Yet, performance variability, orchestration complexity, and workload-specific limitations continue to pose challenges. Future research focused on comprehensive, real-world benchmarking, integration of emerging networking and hardware technologies, and development of robust consistency and orchestration frameworks will be critical to fully harness the potential and drive broader adoption across diverse industrial domains.

REFERENCES

[1]. H. B. Hassan, S. A. Barakat, and Q. I. Sarhan, “Survey on serverless computing,” 2021.

[2]. V. Gulisano, R. Jim, M. Pati, and U. Polit, “StreamCloud : An Elastic and Scalable Data Streaming System,” pp. 1–15.

[3]. F. Li, “Cloud-Native Database Systems at Alibaba : Opportunities and Challenges,” pp. 2263–2272, 2018.

[4]. K. P. Satamraju, “Proof of Concept of Scalable Integration of Internet of Things and Blockchain in Healthcare,” 2020.

[5]. A. R. Gilal, A. A. L. I. Wagan, A. A. L. I. Laghari, I. A. Aziz, and B. A. L. I. Talpur, “BIoMT : A State-of-the-Art Consortium Serverless Network Architecture for Healthcare System Using Blockchain Smart Contracts,” IEEE Access, vol. 10, no. August, pp. 78887–78898, 2022, doi: 10.1109/ACCESS.2022.3194195.

[6]. Q. Zhang et al., “Understanding the Effect of Data Center Resource Disaggregation on Production DBMSs,” vol. 13, no. 9.

[7]. R. Khare, “Serverless Lakehouse Architectures : Beyond the Hype,” vol. 13, no. 46, pp. 20–33, 2025.

[8]. Q. Zhang, Y. Cai, S. Angel, A. Chen, V. Liu, and B. T. Loo, “Rethinking Data Management Systems for Disaggregated Data Centers”.

[9]. H. Bian, D. Geng, P. Guo, Y. Chai, and A. Ailamaki, “Serverless Query Processing with Flexible Performance SLAs and Prices,” no. 1.

[10]. W. Cao et al., “PolarDB Serverless : A Cloud Native Database for Disaggregated Data Centers,” 2021, doi: 10.1145/3448016.3457560.

[11]. R. Toorpu, “Performance Impact on Databases Using Serverless Architectures : An Empirical Study,” pp. 1–4.

[12]. “O P T I M I Z I N G S E RV E R L E S S A R C H I T E C T U R E S F O R,” no. March, 2024.

[13]. R. Mahapatra, S. Ghodrati, B. Hoon, A. Sean, and K. S. Wang, “In-Storage Domain-Specific Acceleration for Serverless Computing,” pp. 530–548, 2024, doi: 10.1145/3620665.3640413.

[14]. T. Bodner, T. Radig, D. Justen, and D. Ritter, An Empirical Evaluation of Serverless Cloud Infrastructure for Large-Scale Data Processing, vol. 1, no. 1. Association for Computing Machinery.

[15]. M. Rodriguez-martinez, “ArcaDB : A Container-based Disaggregated Query Engine for Heterogenous Computational Environments”.

[16]. Y. Su, Y. Sun, M. Zhang, and J. Wang, “Vexless : A Serverless Vector Data Management System Using,” Proc. ACM Manag. Data, vol. 2, no. 3 (SIGMOD), 2024, doi: 10.1145/3654990.

[17]. M. Perron, R. C. Fernandez, D. Dewitt, and S. Madden, “Starling : A Scalable Query Engine on Cloud Functions,” pp. 131–141, 2020, doi: 10.1145/3318464.3380609.

[18]. R. Wang, The Case for Distributed Shared-Memory Databases with RDMA-Enabled Memory Disaggregation [Vision Paper], vol. 1, no. 1. Association for Computing Machinery.
[19]. Y. Zhang et al., "Towards a Shared-Storage-Based Serverless Database Achieving Seamless Scale-Up and Read Scale-Out," in 2024 IEEE 40th International Conference on Data Engineering (ICDE), 2024, pp. 5119–5131. doi: 10.1109/ICDE60146.2024.00384.
[20]. X. Yue, S. Yang, L. Zhu, S. Trajanovski, F. Li, and X. Fu, "Exploiting Wide-Area Resource Elasticity With Fine-Grained Orchestration for Serverless Analytics," IEEE Trans. Netw., vol. 33, no. 1, pp. 398–413, 2025, doi: 10.1109/TNET.2024.3486788.