

GaussDB: Driving Innovation in Cloud-Native Database Technologies

Ronen Grosman, Tianqing Wang, Puya Memarzia

Huawei Canada

{ronen.grosman, wangtianqing2, puya.memarzia1}@huawei.com

1 INTRODUCTION

GaussDB is an enterprise-grade distributed relational database developed by Huawei. With its world-leading shared data disaggregated architecture, versatile distributed storage engine, and broad application compatibility, GaussDB serves thousands of major clients globally and supports the business operations of hundreds of Fortune 500 companies.

2 INNOVATIONS

GaussDB has evolved through long-term product development, extensive user feedback, and continuous innovation, into a world class cloud-native database system [3]. We highlight several key GaussDB innovations below.

Disaggregated Three Layer Architecture: As shown in Figure 1, GaussDB’s architecture shares and disaggregates compute, memory, and storage resources. This allows for independent, on-demand scaling of hardware resources. GaussDB’s architecture is multiwriter, which enhances scalability and throughput by allowing nodes to write data concurrently. This architecture ensures that GaussDB can handle millions of transactions per second (TPS) [2], and efficiently manage resources between different workloads.

Versatile Workload Support: GaussDB is suitable for a variety of different workloads, including OLTP, OLAP, and HTAP. This is achieved through a variety of storage engines such as Columnar, in-place update, and in-memory engine, as well as efficient distributed query processing algorithms. This multi-model support ensures that GaussDB can handle diverse data types and workloads, making it adaptable to various industry needs.

Shared-data Architecture: GaussDB’s shared-data architecture facilitates seamless elasticity, and flexible resource utilization. As clusters grow larger, obtaining optimal performance becomes challenging due to cross-node communication. We employ innovative techniques to minimize this issue, including a novel near-zero overhead transaction management protocol based on synchronized clocks, and extensive data layer enhancements to intelligently partition, affinitize, and route data access.

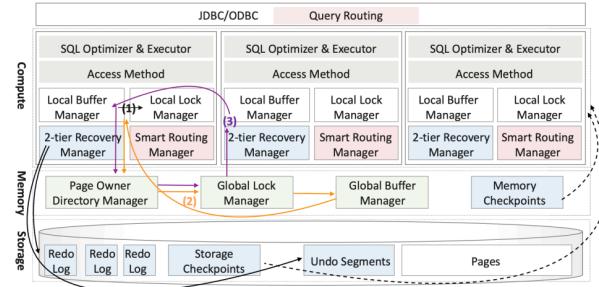


Figure 1: GaussDB Architecture

3 FUTURE CHALLENGES

GaussDB is poised to tackle upcoming challenges and open problems [1]. Key areas of focus include:

- Next generation HTAP and storage engine designs
- Latency hiding for disaggregated architecture
- Scalability both inside compute nodes, and across compute nodes at AZ, and Region level.
- Utilizing heterogeneous hardware (e.g. GPU, DPU)

References

- [1] H. Dong, C. Zhang, G. Li, and H. Zhang. Cloud-native databases: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 2024.
- [2] Guoliang Li, Wengang Tian, Jinyu Zhang, Ronen Grosman, and et al. Gaussdb: A cloud-native multi-primary database with compute-memory-storage disaggregation. *PVLDB*, 17:3786 – 3798, 2024.
- [3] Guoliang Li, Xuanhe Zhou, Ji Sun, Xiang Yu, Yue Han, Lianyuan Jin, Wenbo Li, Tianqing Wang, and Shifu Li. opengauss: An autonomous database system. *Proc. VLDB Endow.*, 14:3028–3041, 2021.