Composability and Interoperability for Federated Data Systems

Haralampos Gavriilidis[★◊], Leonhard Rose[★◊], Joel Ziegler[◊], Jonathan Gerloff[◊], Benedikt Didrich[◊], Midhun Kaippillil Venugopalan[◊], Kaustubh Beedkar[§], Matthias Boehm[★◊], Volker Markl[★◊△]
[★]BIFOLD, [◊]Technische Universität Berlin, [§]Indian Institute of Technology Delhi, [△]DFKI GmbH

ABSTRACT

Modern data systems increasingly adopt composability as a design goal. Prior work on vertical composability has focused on assembling systems from modular internal components, such as parsers, optimizers, and execution engines. In contrast, horizontal composability enables the composition of existing, autonomous systems for ad-hoc data processing pipelines. In this talk, we focus on horizontal composability in federated settings. We present XDB, a system for in-situ cross-DBMS query processing that delegates execution to underlying DBMSs, and XDBC, a high-performance data transfer framework that enables efficient interoperability across systems. Together, they provide a practical foundation for building composable and interoperable federated data systems.

VLDB Workshop Reference Format:
Composability and Interoperability for Federated Data Systems. Third International Workshop on Composable Data Management Systems

TALK SUMMARY

System composability has gained traction in both database management [2, 8] and machine learning [9] systems research. In database management systems, vertical composability enables modular architectures where components, such as optimizers and execution engines, can be reused [5, 6, 8]. However, many real-world scenarios involve multiple independently operated systems. In these settings, horizontal composability becomes essential, requiring both mechanisms to compose data pipelines across data systems and efficient means for data interoperability. This talk presents a holistic perspective on our previously published systems, XDB [4] and XDBC [3], and how they jointly support horizontal composability and interoperability in federated settings.

Federated system pipelines. We present XDB, a system for in-situ cross-database query processing that decentralizes query execution across multiple heterogeneous DBMSes. In many real-world scenarios, data is stored across different systems, and users need to run ad hoc queries over fresh, distributed data. Instead of relying on a central mediating engine (e.g., Presto or Trino), XDB fully offloads query processing to the data sources, including cross-database operations. This reduces resource usage and data movement while leveraging the native capabilities of each source. XDB shows that SQL/MED [7] is a useful starting point for DBMS interoperability, but further mechanisms are needed to support non-relational systems and diverse workloads.

Efficient data exchange. We present XDBC, a framework for adaptive data transfer that addresses the challenges of moving data across systems and engines in federated and cross-platform pipelines. Modern data architectures span heterogeneous environments, including on-premise, cloud, and edge deployments, each with different performance characteristics and constraints. XDBC enables efficient interoperability by decomposing data exchange into modular stages, including reading/writing, de/serialization, de/compression, and sending/receiving, that can be tuned to the underlying environment. While modern formats such as Apache Arrow [1] and its ecosystem (e.g., Arrow Flight, ADBC) provide a strong foundation for efficient interoperability, XDBC offers a more holistic framework that goes beyond formats and protocols to enable adaptable, performance-aware data exchange across diverse environments. Our experience with XDBC shows that robust data interoperability requires not only functional connectors, but also tunable, adaptive mechanisms that handle system diversity.

Toward interoperable system compositions. This talk aims to spark discussion around the practical and architectural challenges of composing data systems in federated settings. We highlight the need for richer interoperability mechanisms that go beyond integration to address performance, tunability, and system diversity. As data infrastructures continue to evolve toward heterogeneity, we believe efficiency-aware, composable interoperability will be essential for future data system architectures.

ACKNOWLEDGMENTS

This work was funded by the German Federal Ministry of Education and Research (grants BIFOLD25B, 01IS17052 / SWC PolyDB).

REFERENCES

[1] 2025. Apache Arrow. https://arrow.apache.org/.
[2] Kaustubh Beedkar, Bertty Contreras-Rojas, Haralampos Gavriilidis, et al. 2023. Apache Wayang: A Unified Data Analytics Framework. *SIGMOD Rec.* 52, 3 (2023), 30–35. https://doi.org/10.1145/3631504.3631510
[3] Haralampos Gavriilidis, Kaustubh Beedkar, Matthias Boehm, and Volker Markl. 2025. Fast and Scalable Data Transfer Across Data Systems. *Proceedings of the ACM on Management of Data* 3, 3 (2025), 1–28. https://doi.org/10.1145/3725294
[4] Haralampos Gavriilidis, Kaustubh Beedkar, Jorge-Arnulfo Quiané-Ruiz, and Volker Markl. 2023. In-Situ Cross-Database Query Processing. In *ICDE*. https://doi.org/10.1109/ICDE55515.2023.00214
[5] Haralampos Gavriilidis, Lennart Behme, Christian Munz, Varun Pandey, and Volker Markl. 2025. CompoDB: A Demonstration of Modular Data Systems in Practice. (2025), 1094–1097. https://doi.org/10.48786/EDBT.2025.97
[6] Haralampos Gavriilidis, Lennart Behme, Sokratis Papadopoulos, et al. 2022. Towards a Modular Data Management System Framework. In *CDMS@VLDB*. https://cdmsworkshop.github.io/2022/Proceedings/ShortPapers/Paper2_HaralamposGavriilidis.pdf.
[7] Jim Melton, Jan Eike Michels, Vanja Josifovski, Krishna Kulkarni, and Peter Schwarz. 2002. SQL/MED: a status report. *ACM SIGMOD Record* 31, 3 (2002).
[8] Pedro Pedreira, Orri Erling, Konstantinos Karanasos, et al. 2023. The Composable Data Management System Manifesto. *PVLDB* 16, 10 (2023), 2679–2685. https://doi.org/10.14778/3603581.3603604
[9] Mark Zhao, Emanuel Adamiak, and Christos Kozyrakis. 2024. cedar: Optimized and Unified Machine Learning Input Data Pipelines. *PVLDB* 18, 2 (2024), 15. https://doi.org/10.14778/3705829.3705861

Proceedings of the VLDB Endowment. ISSN 2150-8097.