Bringing the Operational and Analytical Worlds Together with Lakebase

Matei Zaharia
UC Berkeley and Databricks
matei@berkeley.edu

ABSTRACT

As database workloads increasingly move into large shared-nothing cloud datacenters, the bits storing operational data, analytical tables, streams, etc all sit together on the same disks in the cloud. This creates new opportunities to unify the capabilities of operational and analytical systems, while being mindful of "one size fits all" pitfalls. I'll discuss how Databricks and Neon are exploring this opportunity with Lakebase, an architecture for OLTP DBMSes that leverages open formats and cloud object stores to also enable efficient analytics on the same data and easy interop between the two worlds. Furthermore, since it wouldn't be a 2025 keynote without AI, I'll explain how we are seeing agents change the demand on both types of systems and appear to be resulting in more "analytics-like" workloads on OLTP databases and more "OLTP-like" workloads on analytical ones, primarily by issuing much larger numbers of small exploratory queries. These trends create many exciting new challenges for the research community.

PVLDB Reference Format:
Matei Zaharia. Bringing the Operational and Analytical Worlds Together with Lakebase. PVLDB, 18(12): 5539 - 5539, 2025.

doi:10.14778/3750601.3760524

BIOGRAPHY

Matei is the CTO and co-founder of Databricks and an Associate Professor of Computer Science at UC Berkeley. He started the Apache Spark project during his Ph.D. program at UC Berkeley in 2009 and has worked on other widely used data and AI software, including MLflow, Delta Lake, Unity Catalog and DSPy. His most recent research covers improved cloud infrastructure and new programming models and optimization methods for AI. Matei's research was recognized through the 2014 ACM Doctoral Dissertation Award and the U.S. Presidential Early Career Award for Scientists and Engineers (PECASE).

Proceedings of the VLDB Endowment, Vol. 18, No. 12 ISSN 2150-8097.
doi:10.14778/3750601.3760524