The Three Golden Ages of Database Engineering: From SIGMOD'85 to the Agentic Era

Reynold Xin
Databricks Inc
San Francisco, USA
rxin@databricks.com

ABSTRACT

In 2012, a year before the founding of Databricks, I stumbled upon a book of SIGMOD'85 proceedings in the Berkeley AMPLab office. I felt disheartened, as many of those papers could have been republished in 2012 by simply replacing "megabytes" with "terabytes". Every topic, perhaps with the exception of columnar databases, had already been studied.

It turned out this perspective couldn't have been more wrong. While that early relational era was the first golden age of database engineering, the last decade became the second. "Big data", data science, and cloud computing led to the rise of new workloads on analytic data systems beyond the traditional rollups and drilldowns data warehouses were designed for. This ultimately led the industry to standardize on the Lakehouse, which marries the best ideas of data warehouses with the physical and logical separation of storage and compute to support these new workloads.

In this talk, I will first look back and share how my perspective on database technology at scale has evolved. By examining real-world workloads observed on the Databricks platform, I will discuss how these workloads impact the reality of database engineering, highlighting challenges in scaling a modern data platform that are often overlooked by database systems research and even database vendors.

I will then look forward and discuss why we are entering the third golden age of database engineering. As AI agents change software engineering life cycles, they are introducing not only new workloads (operational analytics, vector search) but also new usage patterns on preexisting OLTP workloads (highly concurrent, iterative, rapid experimentation). They are stressing performance and scalability dimensions that traditional database systems ignore, forcing databases to fundamentally change.

To address this, I will describe the Lakebase architecture, which applies a similar philosophy of separating storage and compute to OLTP databases, serving as a foundation for databases in the agentic world. Finally, I will demonstrate how unifying Lakebase and Lakehouse creates LTAP (Lake Transactional Analytical Processing). By sharing an open storage layer while physically isolating compute, LTAP offers a pragmatic alternative to HTAP, side-stepping the historical trade-offs of combining analytics and OLTP in a single system.

I will conclude by outlining what this third golden age implies for the future of database systems, introducing new fundamental research challenges that open an exciting frontier for the database community. Unlike my 2012 self who feared the field was already solved, I have never been more energized about what lies ahead.

PVLDB Reference Format:
Reynold Xin. The Three Golden Ages of Database Engineering: From SIGMOD'85 to the Agentic Era. PVLDB, 19(12): 4981 - 4981, 2026.
doi:10.14778/3827998.3838713

BIOGRAPHY

Reynold Xin is a co-founder and Chief Architect at Databricks. He leads the engineering teams building systems from Lakehouse to Lakebase, many of which have become industry standards and set world records in benchmarks like TPC-DS, GraySort, and CloudSort. A champion of open data ecosystems, Reynold is a core contributor to foundational open-source projects including Apache Spark, Apache Arrow, Apache Parquet, Apache Iceberg, Delta Lake, and Unity Catalog. He holds a PhD in Computer Science from UC Berkeley, where he focused on large-scale data systems.

This work is licensed under the Creative Commons BY-NC-ND 4.0 International License. Visit https://creativecommons.org/licenses/by-nc-nd/4.0/ to view a copy of this license. For any use beyond those covered by this license, obtain permission by emailing info@vldb.org. Copyright is held by the owner/author(s). Publication rights licensed to the VLDB Endowment.
Proceedings of the VLDB Endowment, Vol. 19, No. 12 ISSN 2150-8097.
doi:10.14778/3827998.3838713