SAP HANA Cloud: Modern Data Management Architecture for Data Lakehouse

Daniel Farrar

SAP Labs Canada, Waterloo, Ontario
dan.farrar@sap.com

At SAP, we are witnessing—and actively shaping—a profound convergence between the traditional relational database paradigm and the modern files-based ecosystem. Historically, files served as raw data containers, exemplified by simple CSV formats, offering openness and durability but lacking the transactional and structural sophistication of relational systems. Over time, this landscape has evolved dramatically. Columnar formats such as Apache Parquet introduced efficient storage and query capabilities, narrowing the gap with analytical databases. The next leap came with formats like Delta Lake, which incorporate Multi-Version Concurrency Control (MVCC), enabling transactional consistency and schema evolution directly within file-based systems. This progression reflects a broader trend: files and their associated compute engines, notably Apache Spark, are assuming capabilities once exclusive to RDBMS while preserving their inherent strengths—extreme durability, open standards, and true separation of compute and storage. By decoupling blob storage from compute, organizations can scale analytics elastically, serving massive numbers of engines without compromising reliability. Equally important, this model shifts the operational and cost burden of analytics from data producers to consumers, safeguarding transactional systems while empowering analytical workloads with unprecedented flexibility. At SAP, we embrace this hybrid approach as the foundation for next-generation cross-application analytics. Our strategy leverages the robustness of relational systems alongside the openness and scalability of file-based architectures. This enables us to deliver rich analytical capabilities across diverse applications without imposing rigid infrastructure constraints. By integrating transactional consistency, schema governance, and high-performance query execution into open file formats, we create an environment where enterprise data remains durable, accessible, and ready for advanced insights—whether through traditional SQL engines or modern distributed frameworks. We believe this convergence is not merely an incremental improvement but a paradigm shift. It redefines how enterprises manage, share, and analyze data at scale, combining the best of both worlds: the reliability and maturity of RDBMS with the flexibility and openness of file-based ecosystems. In this talk, we will explore the architectural principles, design trade-offs, and practical implications of this evolution, illustrating how SAP is driving innovation at the intersection of these two domains.