

From MongoDB to SQL: Efficient and Scalable Data Management with Tiled

Thursday 6 November 2025 17:20 (20 minutes)

Tiled is a general-purpose data access service that unifies heterogeneous scientific data stores behind a structured interface. By mapping diverse storage backends (CSV, HDF5, TIFF, Zarr, Parquet, relational and NoSQL databases) onto a concise set of logical abstractions—tables, arrays, and hierarchical containers—Tiled hides backend details while enabling efficient, sliceable, chunked access. Its HTTP-based architecture supports deployment as a public or private service, modern authentication, fine-grained authorization, caching, and fast data streaming via WebSockets for real-time acquisition.

Developed in the context of the Bluesky project, Tiled provides first-class support for the Bluesky event document model through the TiledWriter callback. This allows to ingest Bluesky run documents into a Tiled catalog, storing scalar data as tables while registering external binary data (e.g. detector images) via StreamResource/StreamDatum documents aligned to a common time index. TiledWriter includes a RunNormalizer that upgrades legacy schemas, can run asynchronously to avoid disrupting experiments, and buffers data during temporary outages.

We present the architecture and performance of the Bluesky–Tiled integration, emphasizing preprocessing and consolidation before data writing. These steps flatten and reindex streaming documents to accelerate later queries. We compare SQL-backed Tiled catalogs with the legacy NoSQL storage solution, demonstrating the improved support of scalable, low-latency lookup, fast random access, and array slicing. Finally, we share experience migrating existing Bluesky datasets from MongoDB to PostgreSQL, quantifying performance gains on representative beamline use cases.

Authors: ALLAN, Dan (Brookhaven National Laboratory); Dr MATVIYCHUK, Yevgen (Brookhaven National Laboratory – NSLS-II)

Co-authors: MCREYNOLDS, Dylan (Lawrence Berkeley National Lab); BISCHOF, Garrett (Brookhaven National Laboratory); Mr WARE, Joseph (Diamond Light Source); Dr MARULANDA ARIAS, Juan (Brookhaven National Laboratory – NSLS-II); Dr AISHIMA, Jun (Brookhaven National Laboratory – NSLS-II); Ms BARRY, Kari (Brookhaven National Laboratory – NSLS-II); MAYTAN, Nate (NSLS2); Dr KARAKUZU, Seher (Brookhaven National Laboratory – NSLS-II)

Presenter: Dr MATVIYCHUK, Yevgen (Brookhaven National Laboratory – NSLS-II)

Session Classification: Community Talks