Konstantin Tarasov

Last updated: 2026-01-14

kosta.tarasov@protonmail.com | github.com/sdf-jkl

Skills

Languages: Rust, Python, SQL
Systems & Storage: Apache Arrow, Parquet, DataFusion
Database Internals: Query optimization, Nested data types, Columnar data formats

Open Source Contributions

Apache DataFusion – Rust (Query Engine Internals)

- Implemented user-defined functions for extracting VARIANT SQL schemas, enabling structured introspection of semi-structured data, by designing expression evaluation logic and type inference from scratch #24
- Designed and implemented a logical query optimization rule, reducing unnecessary scans and improving query performance, by enabling predicate pushdown across UDF boundaries #18789
- Extended UDF execution semantics, fixing correctness issues and improving maintainability, by adding flattening support for nested List(LargeList) types and comprehensive unit tests #18363

Apache Arrow – Rust (Columnar Memory & Execution)

- Implemented page skipping for bitmask-based row filters, improving columnar scan efficiency and query performance by 2%, by pruning irrelevant data pages during execution #9118
- Added bidirectional casts to and from VARIANT, expanding supported query semantics, by extending Arrow's type system, compute kernels, and validation logic #8085 #8201 #8600

MapLibre/Martin – Rust, Bash, SQLite

- Replaced static .mbtiles artifacts, improving security and operational flexibility, by dynamically generating tiles from SQL-backed data sources #1868 #2380
- Increased system reliability, raising confidence in production changes, by overhauling unit and integration tests and ensuring GitHub Actions CI/CD compatibility

Experience

Software Engineer, Metropolitan Transportation Authority (MTA) – New York, NY — Sept 2024 – Present

- Reduced redundant rail closures and manual labor, improving operational efficiency, by developing Airflow- and PySpark-based pipelines syncing data lake records to ArcGIS Enterprise
- Increased data processing throughput 7.5× (200 → 1,500 rows/min), reducing pipeline latency, by building a multithreaded batch-processing engine in Python using Polars
- Improved service response times, reducing time-to-diagnosis for data issues, by building internal database observability tooling to surface query latency and pipeline bottlenecks

Projects

Ray Tracing in Rust — Apr – May 2025

- Built a ray tracing engine supporting lighting, shadows, reflections, and anti-aliasing, demonstrating ownership of a performance-critical system, by implementing core rendering algorithms and vector math
- Improved rendering performance 8×, reducing frame render time, by introducing data-parallel execution with Rayon and restructuring hot execution paths

Education

Baruch College – MS in Information Systems — Dec 2024
Finance University – BS in Economics — June 2022