

Konstantin Tarasov

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

Skills

Languages: Rust, Python, SQL

Systems & Storage: Apache Arrow, Parquet, DataFusion

Databases: Query optimization, Nested data types, Columnar data formats

Open Source Contributions

Apache DataFusion – Rust

- Implementing user defined functions from scratch to extract Variant SQL schema #24
- Implemented optimization rule to support predicate pruning for user-defined functions, improving query performance #18789
- Implemented flatten udf support for List (LargeList) improving code readability, fixing a bug, and adding unit tests #18363

Apache Arrow – Rust

- Enabled support for multiple data type casts to and from Variant #8085 #8201 #8600

MapLibre/Martin – Rust, Bash, SQLite

- Improved repository security and maintainability by replacing static .mbtiles with tiles dynamically generated from SQL scripts #1868 #2380
- Increased project stability and test coverage by overhauling unit and integration tests while ensuring GitHub Actions CI/CD compatibility

Experience

Software Engineer, Metropolitan Transportation Authority (MTA) – New York, NY

Sept 2024 – Present

- Reduced redundant rail closures and saved hours of labor by developing a data pipeline with Airflow and PySpark to sync Data Lake data to ArcGIS Enterprise Server
- Increased data processing throughput $7.5 \times$ ($200 \rightarrow 1,500$ rows/min) by building a multithreaded Python batch-processing library using Polars
- Improved service response time by building a database observability platform using Apache Airflow

Projects

Ray Tracing in Rust

Apr – May 2025

- Built a ray tracer in Rust with lighting, shadows, reflections, and anti-aliasing support to learn 3D rendering and vector math
- Improved rendering performance $8 \times$ by integrating parallelism with rayon

Education

Baruch College – MS in Information Systems

Dec 2024

Finance University – BS in Economics

June 2022