

Azim Afroozeh

Amsterdam, NL · afroozeh3@gmail.com · +31 6 39 36 46 64 · azimafroozeh.org · [GitHub](#)

Summary

Principal database engineer specializing in OLAP databases, compression, and columnar storage. Creator of [FastLanes](#), a novel columnar file format for modern hardware, and [ALP](#), a lossless floating-point encoding adopted in production systems including DuckDB and Parquet, and proposed for [ClickHouse](#). Strong background in low-level performance engineering, SIMD-driven algorithm and data layout design, vectorized execution, and GPU-aware optimization, with experience contributing to open-source database systems.

Skills

Languages: C++, Rust, CUDA

Experience

Polars

Jan 2026–Present

Principal Database Engineer

- Designing and implementing out-of-core execution for workloads exceeding RAM.

Database Architectures Group, CWI

2024–2025

Postdoctoral Researcher

- Led development of [FastLanes](#), a columnar file format with many data-parallel novel compression schemes designed for efficient access on modern CPUs with wide SIMD registers and GPUs.

Database Architectures Group, CWI

2019–2020

Research Intern

- Built a prototype columnar format as part of [MSc thesis](#); formed the basis for [FastLanes](#).

Projects

FastLanes

2022–Present

SIMD and GPU friendly compression layout and columnar file format.

- Up to 40× faster decoding and ~40% better compression than Parquet.
- Led to significant changes in the [Parquet](#) specification.
- Adopted in [DuckDB](#), [Parquet](#), [Lance Format](#), and [Vortex](#); 650+ GitHub stars.

ALP — Adaptive Lossless Floating-Point Compression

2024

SIMD-friendly lossless compression algorithm for floating-point data.

- Integrated into [DuckDB](#), [Parquet](#), [KuzuDB](#), [Vortex](#); proposed for [ClickHouse](#) ([open PR](#)).
- Best Artifact Award at SIGMOD 2024; 150+ GitHub stars.

Education

PhD, Computer Science (Cum Laude)

2020–2024

[Vrije Universiteit Amsterdam](#) & [CWI](#)

Thesis: [FastLanes: A Next-Gen File Format](#)

Co-supervised 5 MSc students on FastLanes and ALP:

- [ALP](#) — novel lossless float compression; orders of magnitude faster than prior encodings thanks to data-parallel layouts of FastLanes; published at SIGMOD 2024.
- [Column Correlations](#) — multi-column compression exploiting cross-column correlations; 20% better compression on already heavily compressed files on Public BI benchmark.
- [Nested Encodings](#) — nested-aware compression for FastLanes; 16% better compression on average (up to 90%), matching or beating Zstd-compressed Parquet on the [RealNest](#) benchmark we created from real nested data to understand real data patterns.
- [Predicate Pushdown](#) — data-parallel filter evaluation on cascaded encodings; avoids full decompression.
- [G-ALP](#) — extends ALP to GPUs; rethinks how lightweight encodings should work on GPU hardware; ~10× faster than NVIDIA nvCOMP; published at DaMoN 2025.

MSc, Computer Science (Cum Laude)

2018–2020

[Vrije Universiteit Amsterdam](#)

Thesis: [Towards a New File Format for Big Data: SIMD-Friendly Composable Compression](#)

BSc, Computer Engineering

2010–2016

[Shahid Bahonar University of Kerman](#)

Service

Reviewer: SIGMOD 2026, DaMoN, DataFormats Workshop

Selected Publications

- *The FastLanes File Format*. VLDB 2025.
- *G-ALP: Rethinking GPU Decompression of LightWeight Encodings*. DaMoN 2025.
- *ALP: Adaptive Lossless Floating-Point Compression*. SIGMOD 2024 (**Best Artifact Award**).
- *Accelerating GPU Data Processing using FastLanes Compression*. DaMoN 2024.
- *FastLanes Compression Layout: Decoding >100B ints/s with Scalar Code*. VLDB 2023.