

LEONARDO KUFFO

DATA SYSTEMS RESEARCHER

Passionate about Data Systems. I like to make databases go fast!
Currently doing research on data compression algorithms & vector similarity search.



lxkr@cw.nl



lkuffo



Amsterdam, NL.



Spanish
English

SKILLS

- Experience with **Data Systems programming** (contributed core features in the **DuckDB** storage engine)
- Experience using a variety of **Data Systems in production** (MySQL, Redshift, MongoDB, Redis, Kafka)
- Fluent in **C++ and Python**
- Extensive knowledge on data **file formats**, **compression** algorithms and **vector indexing** algorithms.
- Cloud Stack (Google Cloud, AWS, Databricks)

WORK EXPERIENCE

CWI • **DATA SYSTEMS RESEARCHER** Jan 2023 - Present | Amsterdam, NL

- Co-created the **ALP algorithm**: the current state of the art in floating-point data compression. Used by default to compress floats & doubles in DuckDB, KuzuDB, and SpiralDB.
- Implemented ALP in the **DuckDB storage engine (PR)**, where ALP is 2-4x faster and compresses 2x more than previous algorithms.
- Created **PDX**: A vertical data layout for vector embeddings that speed up vector search by 2x-7x.
- **PhD. candidate** under the supervision of Prof. Peter Boncz (Database Architectures Group)

Amazon EU • **DATA ENGINEER INTERN** Jun 2022 - Nov 2022 | Luxembourg, LU

- Engineered an architecture in AWS Lambda to monitor metrics of dozens of Redshift clusters.
- Added batch processing functionalities on the global transportation system ETL pipelines running in EMR using Apache Spark.

Shippify Inc. • **BACK-END & LEAD ENGINEER** Oct 2018 - Oct 2021 | Guayaquil, EC

- **Led the development** of Shippify's logistics automation system to automate the management of vehicle fleets and millions of deliveries.
- **Engineered Shippify's vehicle routing system** infrastructure to scale and support the concurrent creation of thousands of vehicle routes.
- Performed **data modeling** in Shippify's relational schemas to support new functionalities and stakeholders' requirements.

CERN • **MACHINE LEARNING RESEARCH INTERN** Jul 2018 - Sep 2018 | Geneva, CH

- Developed Machine Learning models to find workload similarities based on their CPU performance metrics when running benchmarks. The developed ML models used multivariate time series classification techniques.
- Part of CERN Openlab 2018.

EDUCATION

PHD. CANDIDATE

CWI, Amsterdam, NL.
Database Architectures Group
2023-2027

MSC. IN DATA ENGINEERING

University of Amsterdam & VU
GPA: 8.5
2021-2023

BSC. IN COMPUTER SCIENCE

ESPOL, Guayaquil, EC.
GPA: 8.3
2014-2019

LEONARDO KUFFO

DATA SYSTEMS RESEARCHER

PUBLICATIONS



[Google Scholar Profile](#)

- **SIGMOD'25** | PDX: A Data Layout for Vector Similarity Search. *Accepted.*
- **SIGMOD'24** | ALP: Adaptive Lossless floating-Point Compression. [Available Online](#)
- **IEEE Big Data'18** | Detection of Customer Experience in Social Platforms. [Available Online](#)
- **ICEDEG'18** | Mining Entrepreneurs Psycholinguistic Dimensions from Twitter. [Available Online](#)
- **ICEDEG'17** | Predicting Venezuelan states political election results through Twitter. [Available Online](#)

BEYOND MY EXPERIENCE

- **Google Summer of Code 2017.** Worked with the University of Cambridge Biology Centre on a web application development project.
- **Udemy Professor** with +12K students. Teaching Web Scraping, Data Analysis, and Data Visualization.
- **Youtube Professor** with over 1.5M views. Teaching Data Management and Python programming.
- **Data systems conferences volunteer at** CIDR'23, ICDE'24, CIDR'25.
- **Part of the logistics and organization of** The Dutch Belgian Database Day '24, The Dutch Seminar on Data Systems Design (throughout 2024)
- Coffee hobbyist :)