

# DuckDB

Kristo Raun

Data Engineering 2024 Fall



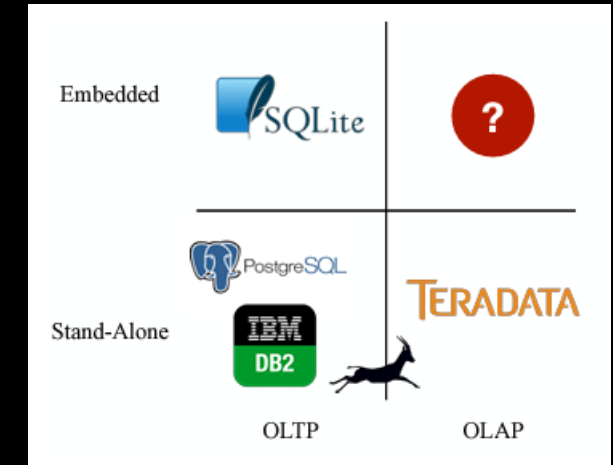
UNIVERSITY OF TARTU

# What is DuckDB?

- In-process SQL OLAP database management system
- Supports a feature-rich SQL dialect complemented with deep integrations into client APIs
- Easy to install, zero external dependencies (C++)

# Why DuckDB?

- No in-process data management solution for OLAP workflows
- Portable, Fast, Extensible
- Columnar-vectorized query execution engine



# Key Features

- Persistence
  - In-memory and saved to disk
- Client APIs
  - Most modern languages supported
- SQL
  - Dialect based on PostgreSQL
- Good support for nested values
  - Array, List, Map, Struct, Union

# Final words...

- DuckDB is great for
  - Complex analytical queries
  - Small to medium datasets
  - Low-overhead PoCs
- DuckDB will not work (or be ideal) if you have
  - $\geq 1$  TB data
  - Many users
  - Streaming data