

# DuckDb



DuckDB

# What is duckdb

- In-process OLAP DBMS
  - “The sqlite for analytics”
- Free and open source (MIT)
- 1.3 million pypi downloads/months
- 12.3k stars on github
- Version 0.9

# Four things to remember about DuckDB

1. It's Fast
2. It's Easy - `pip install duckdb`
3. Works great with python
4. Handles larger than RAM data

# Motivation

- Data processing is an important piece
- Must:
  - Scan different file formats
    - CSV, JSON, Parquet, Postgres, sqlite
  - Integrate with ecosystem tools
    - Plot libraries, ML libraries
  - Be efficient, analytical execution engines
    - Beyond Memory execution
    - Complex query optimisation
  - Support SQL and Relational API

# A new category

In Process



Client-Server



Transactional

Analytical

# The best of database and dataframes

	Dataframes	Client-Server Databases	SQLite	DuckDB
Fast analytical queries	YES	YES	NO	YES
Fast data transfer	YES	NO	NO	YES
Easy to use	YES	NO	YES	YES
Data science integrations	YES	NO	NO	YES
Query Optimisation	SOME	YES	YES	YES
Built in storage	NO	YES	YES	YES
Larger than RAM execution	NO	YES	YES	YES
Relational API	YES	NO	NO	YES
SQL support	NO	YES	YES	YES

# Why you should use it

- Simple to use locally to hack with data and prototype
- Test data warehouse queries locally
- Data Transformations
- Faster/Cheaper performance

# Demo





# Resources

- [chris@justni.com](mailto:chris@justni.com)
- <https://duckdb.org/>
- Active discord
- Pip install duckdb