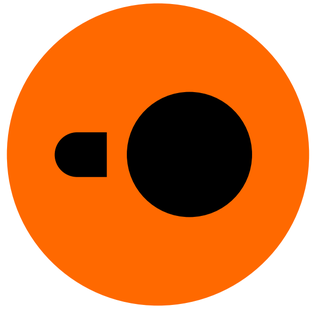


In-Process Analytical Data Management with DuckDB

Hannes Mühleisen

DuckDB Labs

CWI
Centrum Wiskunde & Informatica
CWI

If your data fits in memory there is no advantage to putting it in a database: it will only be **slower** and more **frustrating.**

"Databases are frustrating”

- Installation
- Maintenance
- Data Transfer

- Query Optimization
- Persistence
- Updates & Consistency
- Efficiency & Parallelism

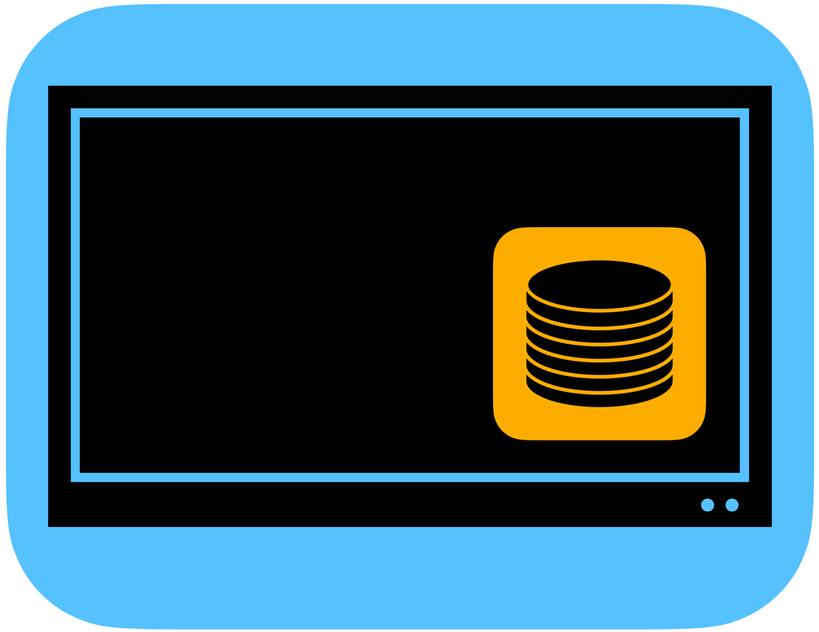
User Experience

DBMS Engine

DuckDB

Simple

In-Process

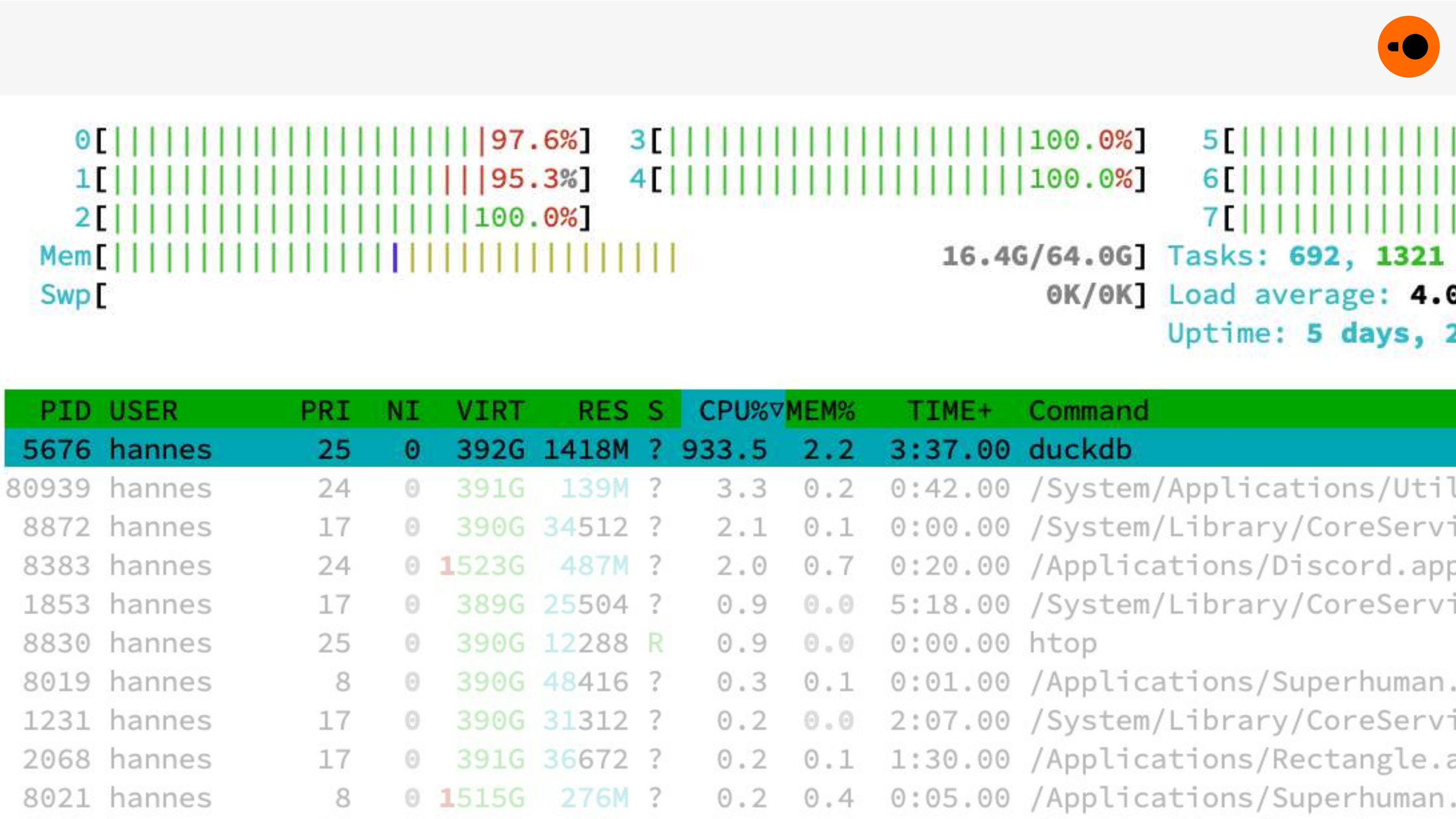


No Dependencies

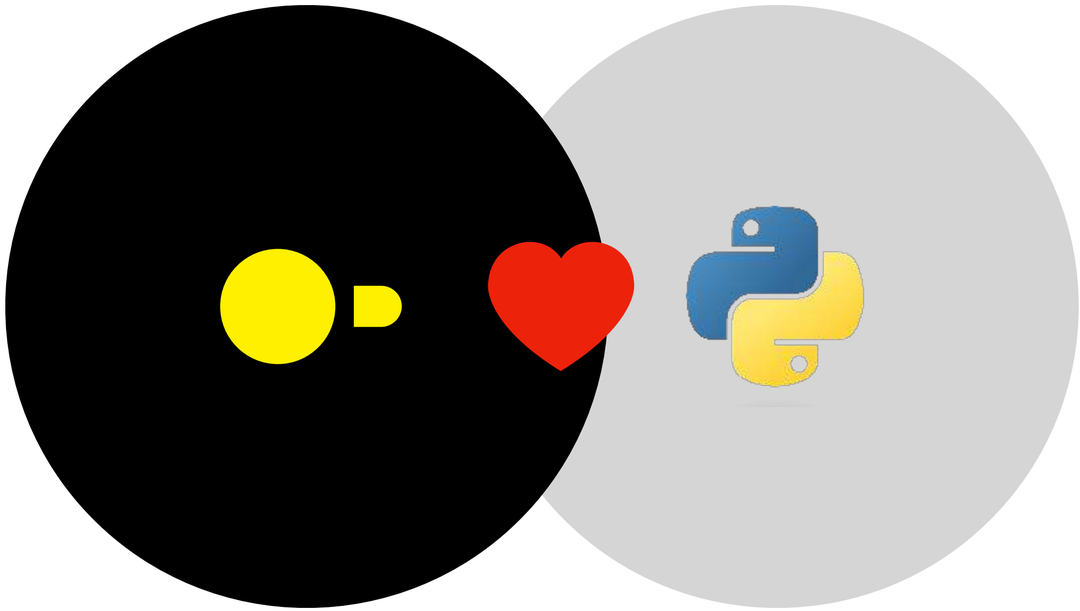
pip install duckdb

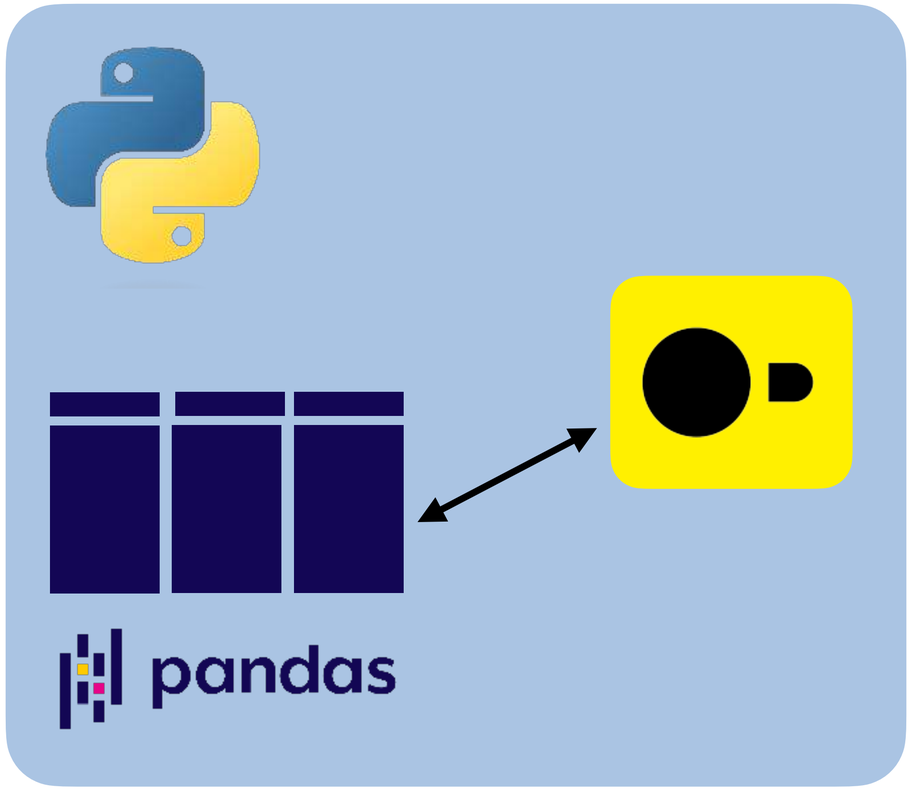
Feature-Rich

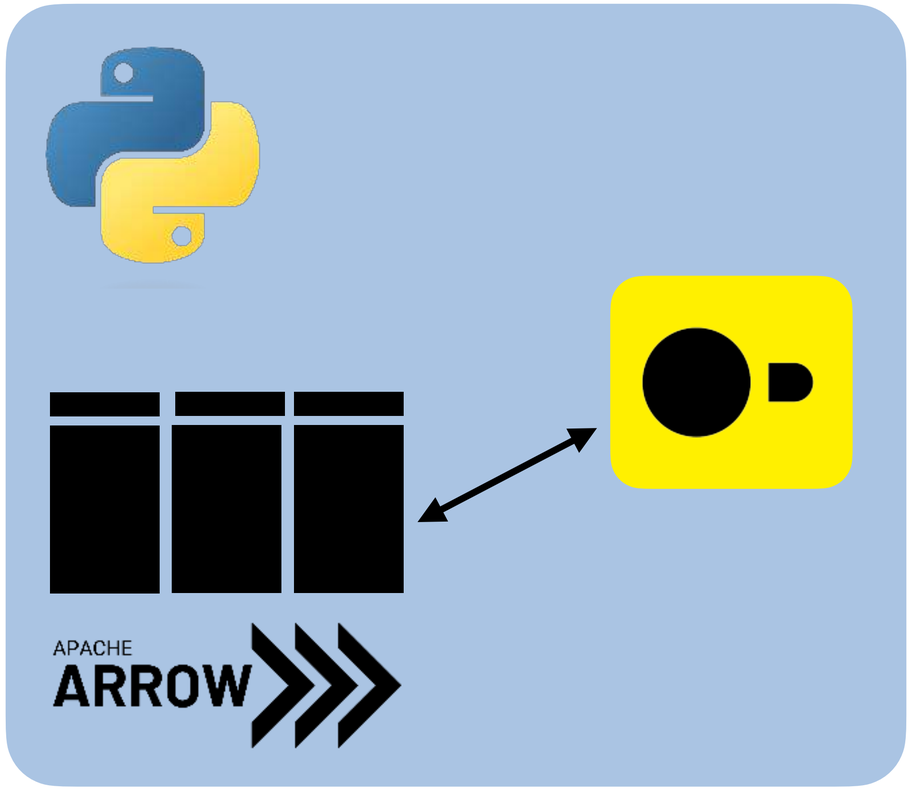
Fast


0[||||||||||||||||||||||||97.6%]
1[||||||||||||||||||||||||95.3%]
2[||||||||||||||||||||||||100.0%]
3[||||||||||||||||||||||||100.0%]
4[||||||||||||||||||||||||100.0%]
5[||||||||||||
6[||||||||||||
7[||||||||||||
Mem[||||||||||||||||||||||||
16.4G/64.0G]
Tasks: 692, 1321
Swp[
0K/0K]
Load average: 4.0
Uptime: 5 days, 2
PID USER PRI NI VIRT RES S CPU%▽MEM% TIME+ Command
5676 hannes 25 0 392G 1418M ? 933.5 2.2 3:37.00 duckdb
80939 hannes 24 0 391G 139M ? 3.3 0.2 0:42.00 /System/Applications/Util
8872 hannes 17 0 390G 34512 ? 2.1 0.1 0:00.00 /System/Library/CoreServi
8383 hannes 24 0 1523G 487M ? 2.0 0.7 0:20.00 /Applications/Discord.app
1853 hannes 17 0 389G 25504 ? 0.9 0.0 5:18.00 /System/Library/CoreServi
8830 hannes 25 0 390G 12288 R 0.9 0.0 0:00.00 htop
8019 hannes 8 0 390G 48416 ? 0.3 0.1 0:01.00 /Applications/Superhuman.
1231 hannes 17 0 390G 31312 ? 0.2 0.0 2:07.00 /System/Library/CoreServi
2068 hannes 17 0 391G 36672 ? 0.2 0.1 1:30.00 /Applications/Rectangle.a
8021 hannes 8 0 1515G 276M ? 0.2 0.4 0:05.00 /Applications/Superhuman.

Free




pandas
Python Process


APACHE
ARROW
Python Process

Demo

Code: https://tinyurl.com/ddbpdams1

One More Thing

20
TRUE TEMPER

“DuckSpark”

Demo

DuckDB

- An in-process analytical data management system
  - Simple, Feature-Rich, Fast, Free. Choose all four.
- Deeply integrated with Python, Pandas, Apache Arrow
- Questions?
- pip install duckdb

duckdb.org

discord.duckdb.org

@duckdb