

PG_DUCKDB: CHALLENGES AND BENEFITS OF ELEPHANTS WITH BEAKS

Jelte Fennema-Nio (@JelteF)

2024-10-25





It's DuckDB inside Postgres

Ehhhmm what???

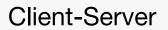




Lightweight in-process SQL Analytics Engine

In-Process







Transactional



What does that look like?

Well... initially like this



Lots of core differences

С	VS	C++
elog(ERROR, …)	VS	exceptions
processes	VS	threads

So we did lots of work



And now it's ready to use duckdb / pg_duckdb Q 🕴 | + 🗸 🖸 II 🗗 <> Code 🕥 Issues 41 1 Pull requests 11 🖓 Discussions Actions . . . Releases / v0.1.0 0 n Convert to draft Compare v0.1.0 Latest 👩 JelteF released this 1 minute ago 🛛 <> v0.1.0 - - 1d00d1b 🕗 Initial release 🎉



What can it do?

- Read/write parquet & csv from/to blob storage
- Use DuckDB engine on Postgres tables
- TEMP tables in DuckDB columnar format
- Offload analytics to MotherDuck

But is it fast???



It depends...

But sometimes yes!



1. Set up TPC-DS with 10GB and no indexes

- 1. Set up TPC-DS with 10GB and no indexes
- 2. Run Q1 -> $\overline{\underline{z}}$ $\overline{\underline{z}}$ wait 10 minutes and give up

- 1. Set up TPC-DS with 10GB and no indexes
- 2. Run Q1 -> <u>z</u> <u>z</u> wait 10 minutes and give up
- 3. SET duckdb.force_execution = true;

- 1. Set up TPC-DS with 10GB and no indexes
- 2. Run Q1 -> $\overline{\underline{X}}$ $\overline{\underline{X}}$ wait 10 minutes and give up
- 3. SET duckdb.force_execution = true;
- 4. Run Q1 -> done in 450ms!

- 1. Set up TPC-DS with 10GB and no indexes
- 2. Run Q1 -> $\overline{\underline{X}}$ $\overline{\underline{X}}$ wait 10 minutes and give up
- 3. SET duckdb.force_execution = true;
- 4. Run Q1 -> done in 450ms!
- 5. Easiest query optimization ever 🎉



Please try it

- MIT licensed
- github.com/duckdb/pg_duckdb
- Feedback welcome
- Blog at motherduck.com/blog

