

Open in app ↗

Medium

 Search

PostgreSQL and DuckDB: Supercharging ad-hoc Data Analysis and ETL (Josef Machytka: Speaker portfolio)



Josef Machytka

2 min read · Feb 9, 2025



Listen



Share



More

(See my bio, other talks in my portfolio and my speaker experience [in the covering article.](#))

Duration: 45 minutes

Target Audience: Data Analysts, Data Scientists, App developers

Overview:

DuckDB is a powerful and practical tool, designed by data analysts in academia to address common use cases efficiently. It was born out of PostgreSQL DNA — it uses PostgreSQL parser and closely follows its SQL syntax. Therefore it efficiently complements PostgreSQL workflows by streamlining ad-hoc data analysis, ETL processes, cross-database queries, and basic data migrations. Its lightweight highly efficient architecture eliminates the need for extensive setup or specialized tools, making it a valuable asset for simplifying complex tasks.

This talk provides actionable insights into integrating DuckDB with PostgreSQL to enhance everyday data workflows. Attendees will explore practical techniques, including the use of DuckDB extensions and Python scripts, to improve agility and productivity in data handling. The focus is on demonstrating DuckDB's ability to deliver simple and powerful solutions for common analytical challenges. Author regularly writes articles on medium.com about abilities of DuckDB and follows its latest development.

Key Takeaways:

- Small Data Manifesto as a growing trend in data analysis
- Understanding of DuckDB's main features, strengths, and limitations
- Usage of DuckDB as a very efficient enhancement tool for PostgreSQL
- Optimization of analytical workload by combining capabilities of PostgreSQL and DuckDB

Recording: [PostgreSQL MeetUp for All 2025.01.08](#)

Slides: [academia.edu](#)

Presented at:

- [Prague PostgreSQL MeetUp October 2024](#)
- NetApp internal tech talk — 2024.11.26
- [PostgreSQL MeetUp for All — 2025.01.08](#)

Reactions:

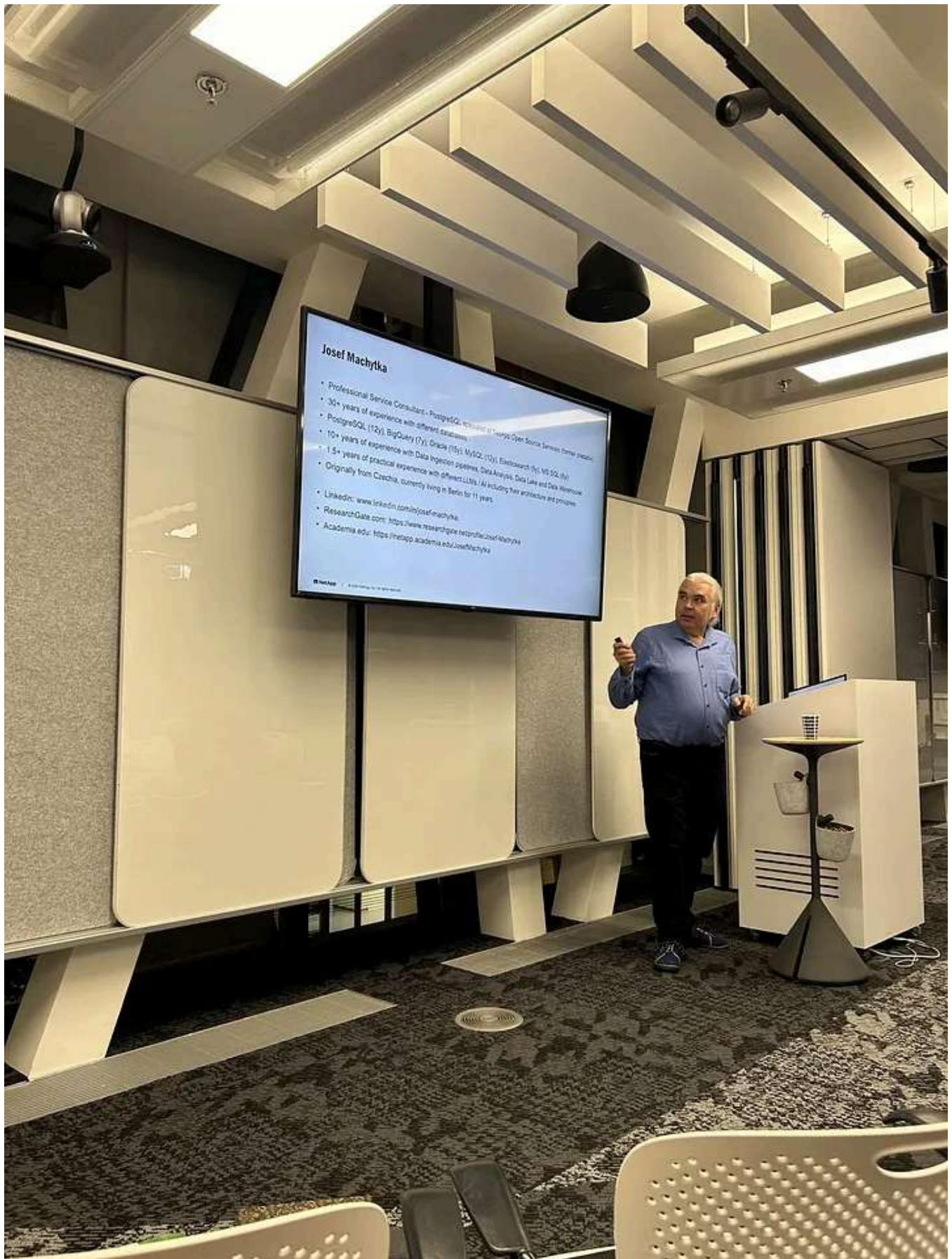
“Thank you so much for your talk and participation in our meetup! We’re very honored and grateful to have you and your expertise. I’ve researched DuckDB a lot in the last year

and you did a fantastic job of covering everything.”

(Elizabeth Garrett Christensen — Crunchy Data)



© Elizabeth Garrett Christensen Crunchy Data — picture from LinkedIn post about the talk



© Igor Gavrilov ([LinkedIn post](#)) — Prague PostgreSQL Meetup October 2024

[Postgresql](#)[Duckdb](#)[Speakers](#)[Speakerhub](#)[Data Analysis](#)



Edit profile

Written by Josef Machytka

68 Followers · 25 Following

I work as PostgreSQL specialist & database reliability engineer at credativ GmbH.

No responses yet



Josef Machytka he/him

What are your thoughts?

More from Josef Machytka



Josef Machytka

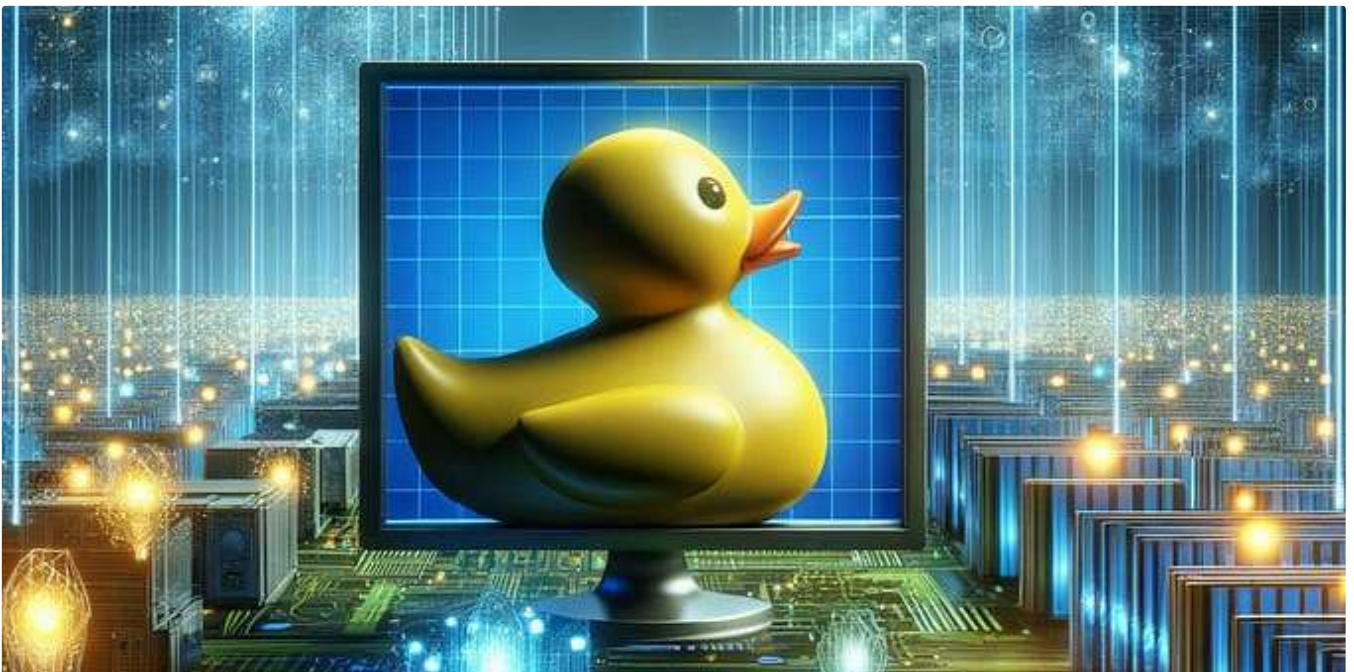
DuckDB Database File as a New Standard for Sharing Data?

This is not my original idea; I came across it in an excellent article titled “DuckDB Beyond the Hype” by Alireza Sadeghi. However, it...

Dec 30, 2024

👏 24

💬 3

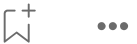


Josef Machytka

Quick and Easy Statistics and Histograms with DuckDB

DuckDB is an exceptional tool that demonstrates how tasks requiring sometimes considerable manual effort in other tools can be accomplished...

Dec 16, 2024 26



 Josef Machytka

PostgreSQL JSONB Operator Classes of GIN Indexes and Their Usage

Throughout 2024, I worked on an internal project exploring the use of JSONB data in PostgreSQL and its various indexing options. During...

Jan 8 5



Bob	2100.0	600.0				
Charlie	2300.0	1500.0	1100.0			

D pivot pg.sales on (product,year) using sum(sales_amount) group by salesperson order by salesperson;

salesperson varchar	(Laptop, 2022) double	(Laptop, 2023) double	(Phone, 2022) double	(Phone, 2023) double	(Tablet, 2022) double	(Tablet, 2023) double
Alice	1200.0	1400.0	800.0	900.0	300.0	400.0
Bob	1000.0	1100.0	600.0			
Charlie	1100.0	1200.0	700.0	800.0	500.0	600.0

D pivot pg.sales on (year,product) using sum(sales_amount) group by salesperson order by salesperson;

salesperson varchar	(2022, Laptop) double	(2022, Phone) double	(2022, Tablet) double	(2023, Laptop) double	(2023, Phone) double	(2023, Tablet) double
Alice	1200.0	800.0	300.0	1400.0	900.0	400.0
Bob	1000.0	600.0		1100.0		
Charlie	1100.0	700.0	500.0	1200.0	800.0	600.0

D pivot pg.sales on (year) using sum(sales_amount) group by salesperson order by salesperson;

--	--	--	--	--	--	--

 Josef Machytka

Easy and Intelligent Pivot Tables with DuckDB

After exploring the various capabilities of DuckDB in my earlier articles, I want to focus more on its powerful data analytical...

Dec 4, 2024  7  1



See all from Josef Machytka

Recommended from Medium

Query Optimization Method	Execution Time	Performance Improvement (vs. No Index)
No Index	42,049 ms (≈42 sec)	Baseline
With B-tree Index	9,684 ms (≈9.7 sec)	77% faster
With Chunk-Skipping Index + Columnstore	304 ms (0.3 sec)	99.28% faster

 In Timescale by Team Timescale

Handling Billions of Rows in PostgreSQL

Here's how to scale PostgreSQL to handle billions of rows using Timescale compression and chunk-skipping indexes.

Jan 17  72  1





 Ramesh Ponnusamy

Query Any Data, Anywhere-Simplify Workflows with DuckDB

Engineering is all about simplicity. We often use lightweight technologies to solve problems instead of overly complex ones, wherever we...

★ Jan 6 🖱 48 💬 1



Lists



Practical Guides to Machine Learning

10 stories · 2215 saves



ChatGPT prompts

51 stories · 2605 saves



Staff picks

819 stories · 1637 saves



Natural Language Processing

1958 stories · 1605 saves

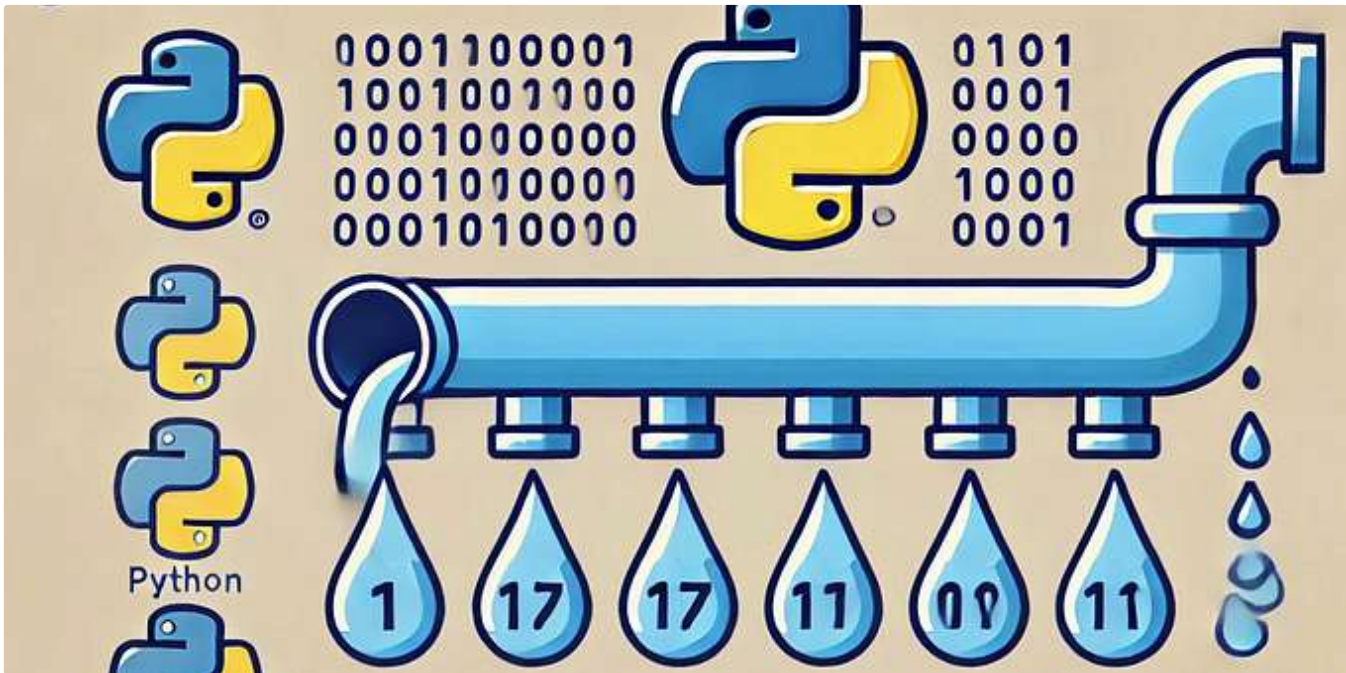


In Towards Dev by Merve Gamze Cinar

Optimizing dbt Models with Common Table Expressions (CTEs)

Writing SQL is one thing, but designing a well-structured dbt model is a whole different game. One key aspect that often separates good...

Feb 19 🖐️ 19 💬 1



In Dev Genius by Ani Palakurthi

10 Essential but Lesser-Known Python Features for ETL Pipelines

Python offers many underrated features and libraries that can supercharge your ETL process.

Feb 7 🖱 52 💬 1



In Top Python Libraries by Ajay Parmar

9 Stunning Scientific Plots I Created Using This Powerful Python Library

Through this article, I aimed to enhance scientific work.



Feb 21 🖱 112 💬 1



BastiaanRudolf

Microsoft Fabric: The Good, The Bad & The Ugly

A few weeks ago, I started a greenfield implementation of Fabric, Microsoft's next-gen data platform that promises unified experiences...

★ Feb 20 🤝 26 💬 1



See more recommendations