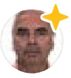


Josef Machytka: Speaker Portfolio — PostreSQL, DuckDB, Databases in General

Josef Machytka
8 min read · Nov 3, 2024

Listen Share More

Expert Talks on PostgreSQL, DuckDB, Databases in general, Data LakeHouse, Data Ingestion and Data Analysis

Introduction

With over 30 years of experience in database technologies, I have cultivated a deep understanding of diverse database systems, including extensive work with PostgreSQL. I worked with multiple heterogeneous data ingestion and data processing pipelines, focusing on delivery of valuable business results. The need to solve often very complicated technical problems equipped me with the expertise necessary for insightful and impactful talks. I am passionate about sharing my knowledge, empowering others in the field of databases and data processing and making complex topics accessible.

My talks are living and evolving, I steadily improve them with new content reflecting progress in new versions of the corresponding software. My speaking style combines practical insights with important technical details, ensuring that audiences walk away with really valuable knowledge they can apply. Slides contain a lot of additional information, making them valuable resources even outside the context of my presentations.

Thank you for considering me as a speaker for your next event. I look forward to the opportunity to share my knowledge and help your audience gain deeper insights into PostgreSQL and database topics.

About Me

Name: Josef Machytka
My current job: Professional Service Consultant, PostgreSQL specialist at NetApp Open Source Systems

Experience:
* 30+ years of production experience with different databases:
PostgreSQL 13 years, BigQuery 7y, Oracle 15y, MySQL 12y,
Elasticsearch 5y, MS SQL 5y, Sybase ASE, FoxPro
* 10+ years of experience with high volume and velocity data ingestion pipelines, Data Analysis, Data Warehouse and Data Lakehouse architecture
* 2+ years of practical experience with different LLMs, their architecture and principles

Accounts: LinkedIn, ResearchGate.net, Academia.edu
Mastodon: @JosefMachytka@me.dm

Current Talk Portfolio

- *GIN, BTREE_GIN, GIST, BTREE_GIST, HASH and BTREE indexes on JSONB data*
- *PostgreSQL and DuckDB: Supercharging ad-hoc Data Analysis and ETL*
- *Building a Data Lakehouse with PostgreSQL: Dive into Formats, Tools, Techniques, and Strategies*
- *PostgreSQL Connections Memory Usage: How Much, Why, and When?*
- *Partitioning and Clustering: An Overview of Solutions with a deep dive into PostgreSQL implementation*

See my latest speaking engagements at the end of this article.

GIN, BTREE_GIN, GIST, BTREE_GIST, HASH and BTREE indexes on JSONB data

Duration: 45 minutes

Target Audience: Application developers, data analysts

Overview: Talk summarizes several months long and still ongoing internal project testing usage and performance of GIN and BTREE_GIN with different operator classes, GIST and BTREE_GIST indexes for GeoJSON data and also standard HASH and BTREE indexes specifically on JSONB data. Tested on several real life datasets with a total size of dozens of GBs. Also, the influence of TOAST compression algorithms, parallelism, memory settings, table statistics on processing JSONB data was tested. Objective of this project was to gather relevant experience to be able to help our customers with their problems, because the majority of articles on the web about JSONB data in PostgreSQL show only trivial examples without any reasonable value for developers solving multiple performance issues related to JSONB data. The talk also discusses practical limitations developers would face if they try to fully decompose JSONB data into relational tables.

Key Takeaways:

- Understanding of use cases and performance of different types of indexes for JSONB data
- The impact of system settings like TOAST compression, parallelism, and memory on performance and usage of indexes on JSONB data
- Limitations and considerations for decomposing JSONB data into relational structures
- Insights into internal structure of different types of indexes

Slides: academia.edu

Presented at:

- Prague PostgreSQL Developer Day 2024 (article on NetApp blog)
- Swiss PG day 2024 (article on NetApp blog)
- Berlin PostgreSQL MeetUp October 2024 (MeetUp entry)

PostgreSQL and DuckDB: Supercharging ad-hoc Data Analysis and ETL

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 complements PostgreSQL workflows by streamlining ad-hoc data analysis, ETL processes, cross-database queries, and basic data migrations. DuckDB's lightweight yet 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.

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)*

Building a Data Lakehouse with PostgreSQL: Dive into Formats, Tools, Techniques, and Strategies

Duration: 45 minutes
Target Audience: Application Developers, Database Administrators, Data Analysts

Overview:

The evolution of Data Warehouses, Data Lakes, and Data Lakehouses has been marked by many buzzwords, fluctuating trends, and tools that often over-promised but under-delivered. While there are numerous materials on these topics, most of them provide mostly introductory overviews and focus narrowly on a single technology. And there are even many different opinions about what exactly is Data Lakehouse.

This talk discusses different ways how to understand this topic. It explores data formats and frameworks like Parquet, Apache Iceberg, Delta Lake, Apache Hudi. Discusses different architectures of Data Lakehouse solutions. Also key challenges will be addressed, such as effective Data Governance, compliance with privacy and security standards, and comprehensive data quality checks.

Last part of the talk address current AI hype with its many promises and proposes realistic overview of real capabilities of current Large Language Models and their use cases in Data Lakehouses.

PostgreSQL is extremely well equipped to play a major role in the current Data Lakehouse and AI boom.

Key Takeaways:

- A comprehensive overview of Data Lakehouse architecture
- Insights into key data formats and frameworks in modern Data Lakehouses
- Practical ideas for implementing Data Governance practices
- Realistic view of real capabilities of current LLMs in scope of Data Lakehouses

Slides: academia.edu

Presented at:

- Prague PostgreSQL Developer Day 2025
- NetApp Internal Talk 2025.01.23

PostgreSQL Connections Memory Usage: How Much, Why, and When?

This talk explores the memory usage of PostgreSQL connections on Debian/Ubuntu running on x86–64 architecture. It gives an overview of memory management concepts, explaining key metrics like virtual, resident, and proportional memory sizes. It also covers various Linux tools for displaying memory usage.

The second part presents practical measurements of PostgreSQL memory usage, based on data from /proc/PID/smaps. It explains why RSS numbers for PostgreSQL connections appear so large after query execution and demonstrates that the actual unique memory usage is only a few dozen megabytes.

Finally, the talk examines how PostgreSQL connections allocate and release additional memory during query processing. It also clarifies where work_mem fits into these numbers and visualizes the process with plots.

Key Takeaways:

- The large RSS values in long-running sessions mostly come from linked shared_buffers
- A newly created connection consumes only up to 10 MB of physical memory, independent of the work_mem setting
- Additional memory is allocated and later released as queries execute
- Work_mem is a "soft maximum limit" — it may not be fully used, but it can also be exceeded

Slides: not available online yet

Presented at:

- NetApp internal talk 2025.01.27

Partitioning and Clustering: An Overview of Solutions with a deep dive into PostgreSQL implementation

Duration: 45 minutes
Target Audience: App developers, system architects

Overview: In this presentation, we will examine the implementation of partitioning and clustering in several database systems, such as BigQuery, Snowflake, Oracle and MySQL. Following that, we will discuss a detailed analysis of PostgreSQL's approach to inheritance, partitioning, and clustering. We will check database parameters that affect performance of these solutions, including rarely used enable_partitionwise_aggregate and enable_partitionwise_join parameters, will compare the results of performance tests between a single large table and partitioned tables on different datasets, look at efficiency of indexes, and discuss the application of multi-level partitioning for different use cases. Additionally, we will delve into details of memory usage, statistics and query optimization. We will also share some tips and tricks — how to subsequently implement partitioning on existing table, how to create child tables with additional columns.

Key Takeaways:

- Understanding of partitioning & clustering across different database systems
- In-depth understanding of PostgreSQL implementation
- Practical tips and tricks for performance optimization and implementation of partitions

Slides: not available online

Presented at:

- NetApp internal workshop 2024.11.12

My latest speaking engagements

1. **2025.01.29 Prague PostgreSQL Developer Day 2025**
   “Building a Data Lakehouse with PostgreSQL: Dive into Formats, Tools, Techniques, and Strategies”
2. **2025.01.27 NetApp internal talk**
   “PostgreSQL Connections Memory Usage: How Much, Why, and When?”
3. **2025.01.23 NetApp internal talk**
   “Building a Data Lakehouse with PostgreSQL: Dive into Formats, Tools, Techniques, and Strategies”
4. **2025.01.08 PostgreSQL MeetUp for All online event**
   “PostgreSQL and DuckDB: Supercharging ad-hoc Data Analysis and ETL”

5. **2024.11.26 NetApp internal talk**
“PostgreSQL and DuckDB: Supercharging ad-hoc Data Analysis and ETL”

6. **2024.11.15 NetApp Tech Talk**
“Many facets of AI hallucinations: factual errors, deep fakes, and creativity”

7. **2024.11.12 Internal Workshop —** 3 talks:
“PostgreSQL 17: New Features”
“PostgreSQL Architecture: How PostgreSQL functions under the Hood”
“PostgreSQL Partitioning: Overview of Solutions, Tips & Tricks, Performance”

8. **2024.10.29 Prague PostgreSQL MeetUp October 2024**
“PostgreSQL and DuckDB”

9. **2024.10.16 NetApp internal workshop —** “AI Workshop: Exploring Artificial Intelligence” — 3 talks:
- “The AI Dilemmas: The Philosophical, Ethical, and Legal Implications of AI”
- “Understanding the Magic of LLMs: A Deep Dive into the Internal Structure of LLMs”
- “Many facets of AI hallucinations: factual errors, deep fakes, and creativity”

10. **2024.10.09 PostgreSQL MeetUp Berlin October 2024**
- “GIN, BTREE_GIN, GIST, BTREE_GIST, HASH and BTREE indexes on JSONB data”

11. **2024.09.27 NetApp internal talk**
- “Beyond the Buzzwords: A pragmatic explanation of LLM terminology”

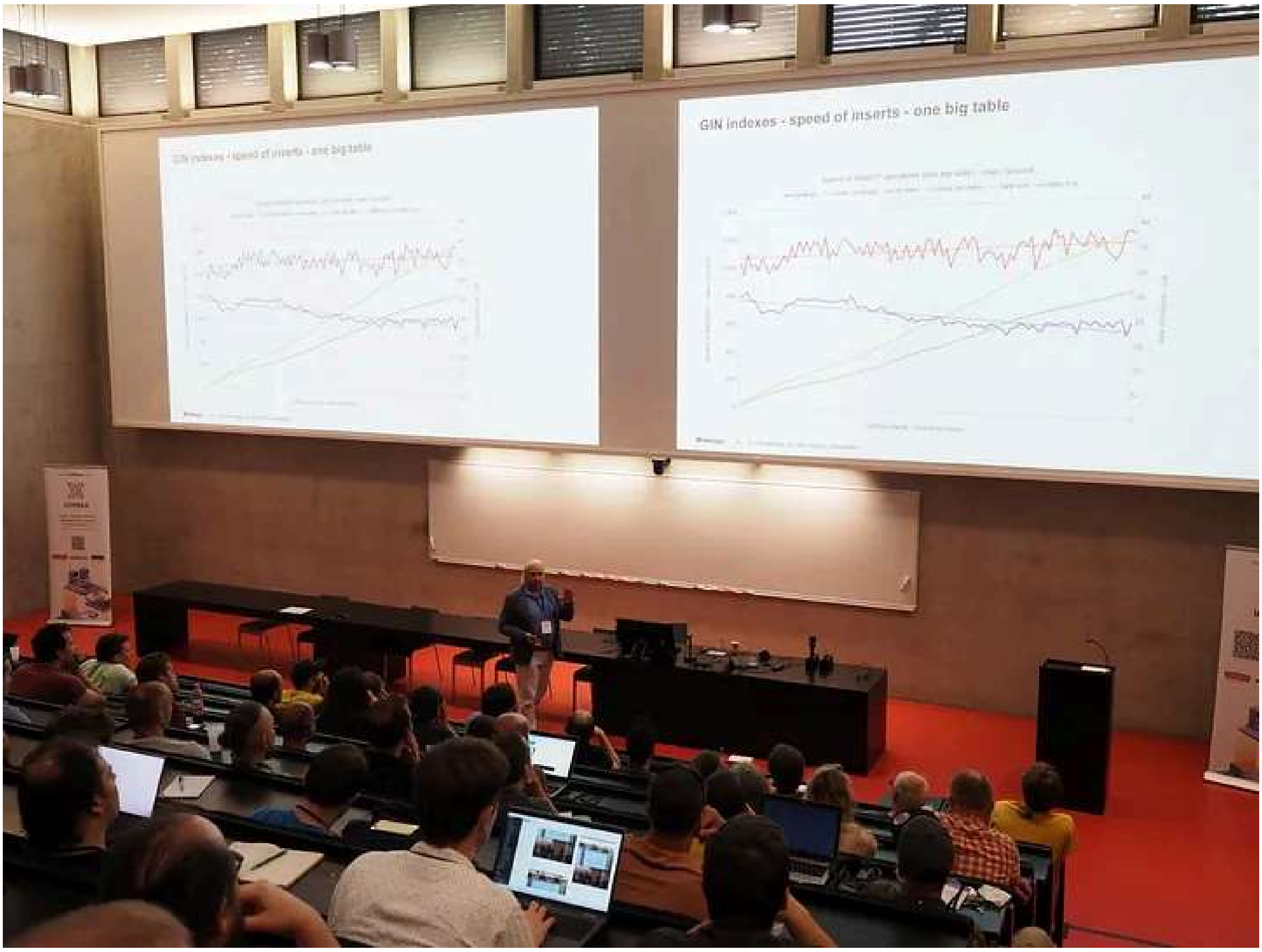
12. **2024.10.13 NetApp internal talk**
- “The Art and Science of AI Prompt Engineering”

13. **2024.07.16 PostgreSQL MeetUp Berlin July 2024**
- Lightning talk: “Can PostgreSQL have a more prominent role in the AI boom?”

14. **2024.06.27 Swiss PG day 2024**
- “GIN, BTREE_GIN, GIST, BTREE_GIST, HASH and BTREE indexes on JSONB data”

15. **2024.06.05 Prague PostgreSQL Developer Day 2024**
- “GIN, BTREE_GIN, GIST, BTREE_GIST, HASH and BTREE indexes on JSONB data”

Photos from my talks

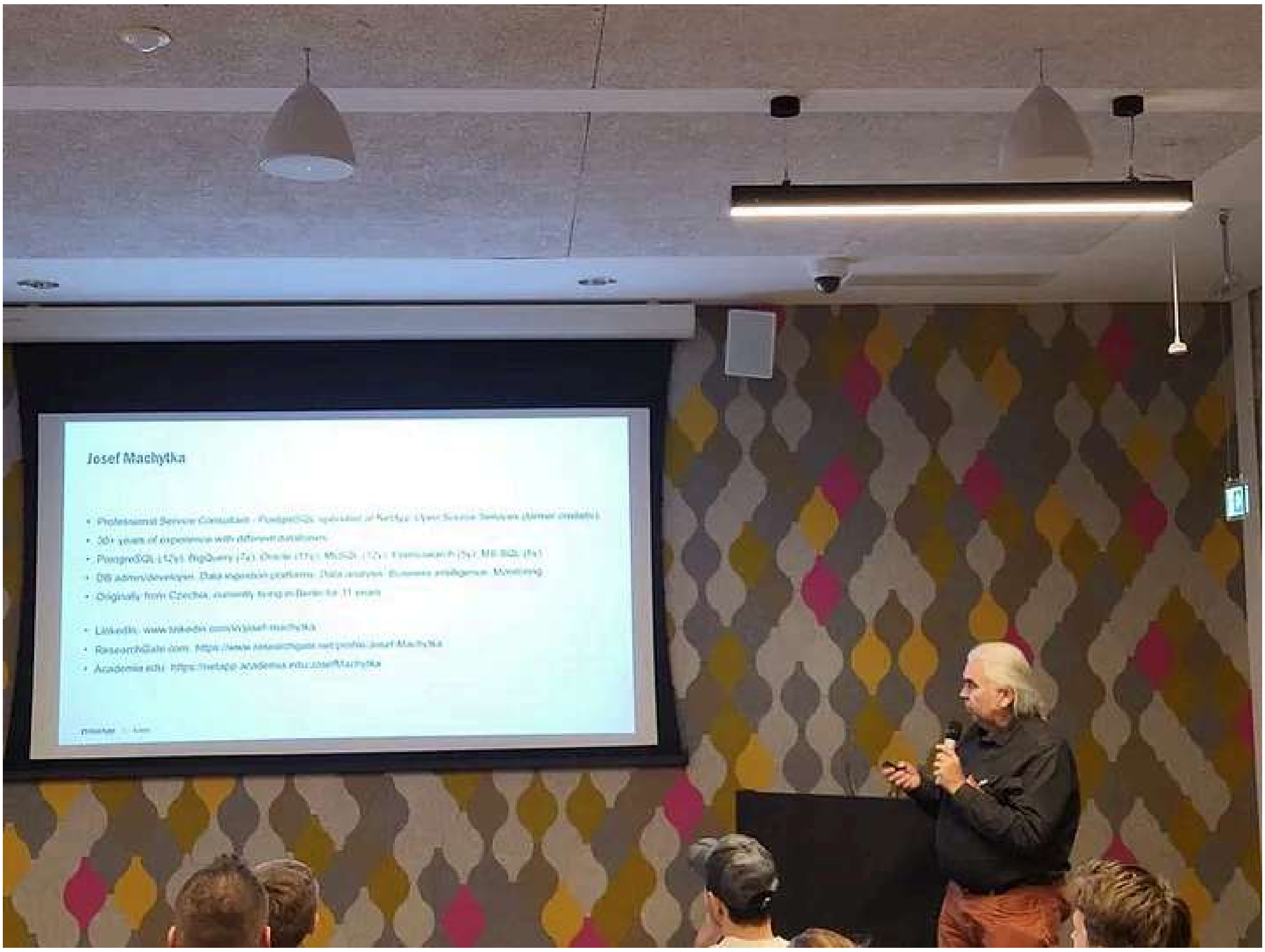


© Tomas Vondra P2D2 — Prague PostgreSQL Developer Day 2024 — talk about indexes on JSONB data

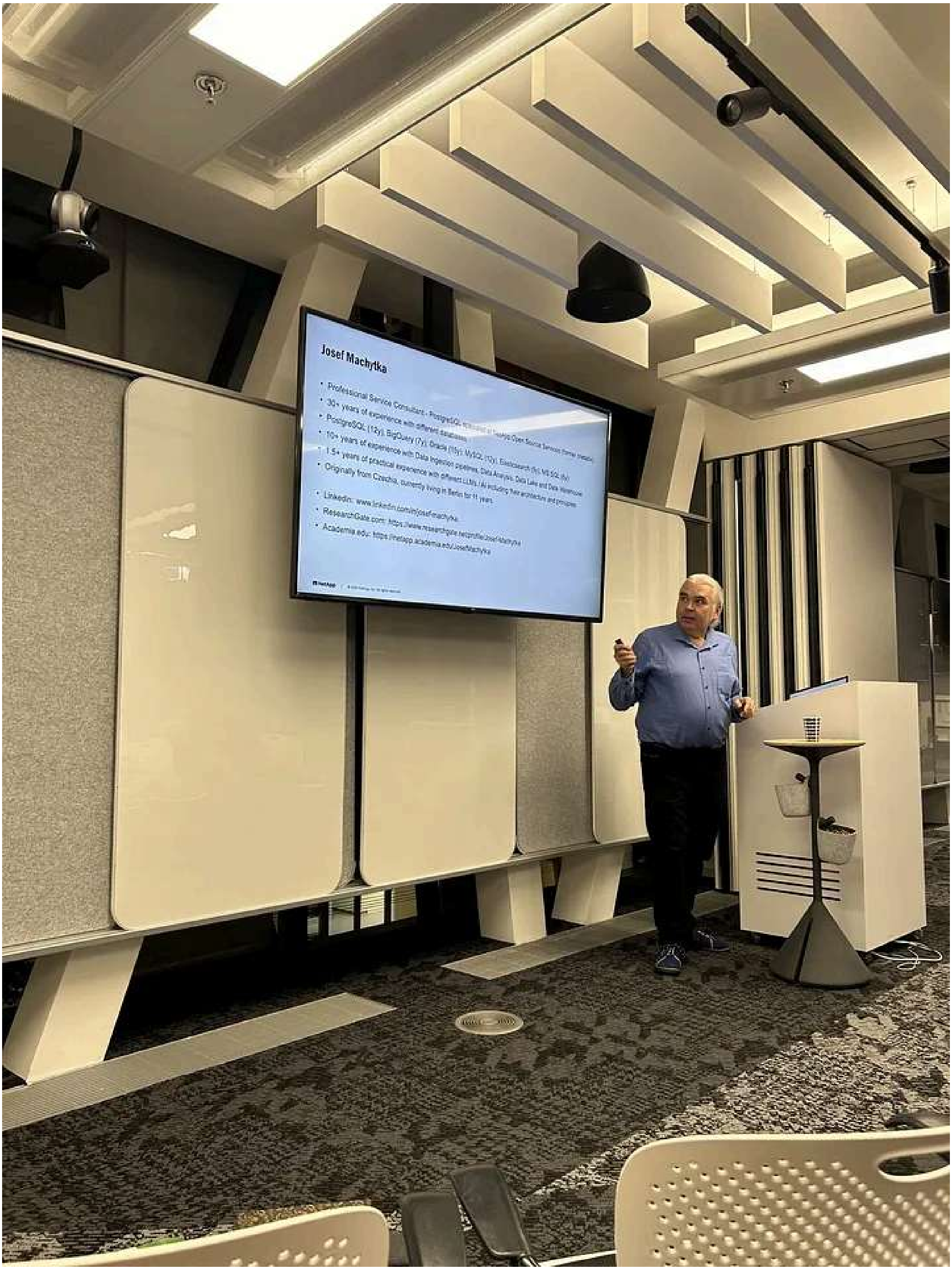
experience with different databases.
(11y), BigQuery (7y), Oracle (15y), MySQL (12y), Elasticsearch (5y), MS SQL (5y).
eveloper, Data ingestion platforms, Data analysis, Business intelligence, Monitoring.
om Czechia, currently living in Berlin for 11 years.
ww.linkedin.com/in/josef-machytka.

© organizers of Swiss PG day — Swiss PG day 2024 — talk about indexes on JSONB data

© Andreas Scherbaum — PostgreSQL Meetup Berlin July 2024 — lightning talk about PostgreSQL and AI



© Andreas Scherbaum — PostgreSQL Meetup Berlin Octobec 2024 — talk about indexes on JSONB data


Josef Machytka
Professional Service Consultant - PostgreSQL [illegible] at NetApp Open Source Services (former Instaclustr)
30+ years of experience with different databases
PostgreSQL (12y), BigQuery (7y), Oracle (15y), MySQL (12y), Elasticsearch (5y), MS SQL (5y)
10+ years of experience with Data ingestion pipelines, Data Analysis, Data Lake and Data Warehouse
1.5+ years of practical experience with different LLMs / AI including their architecture and principles
Originally from Czechia, currently living in Berlin for 11 years
LinkedIn: www.linkedin.com/in/josefmachytka
ResearchGate.com: https://www.researchgate.net/profile/Josef-Machytka
Academia.edu: https://netapp.academia.edu/JosefMachytka
NetApp

© Igor Gavrilov (LinkedIn post) — Prague PostgreSQL Meetup October 2024 — talk about DuckDB

NetApp
Building a Data Lakehouse with PostgreSQL
Dive into Formats, Tools, Techniques, and Strategies
Josef Machytka <josef.machytka@netapp.com>
NetApp Open Source Services / Credativ
2025-01-29 - Prague PostgreSQL Developer Day 2025
xata
Test. Deploy. Scal
Postgres databas
pgroll
Roll changes without dow
pgroll.com
pgstream
Stream changes to other d
stores
xata.io/pgstream
maki
Boost your staging and de
environments
maki.tech
xata.io

© organizers of Prague PostgreSQL Developer Day 2025

© organizers of Prague PostgreSQL Developer Day 2025

Postgresql

Data Analysis

Data Ingestion

Duckdb

Edit profile

Written by Josef Machytka

61 Followers · 22 Following

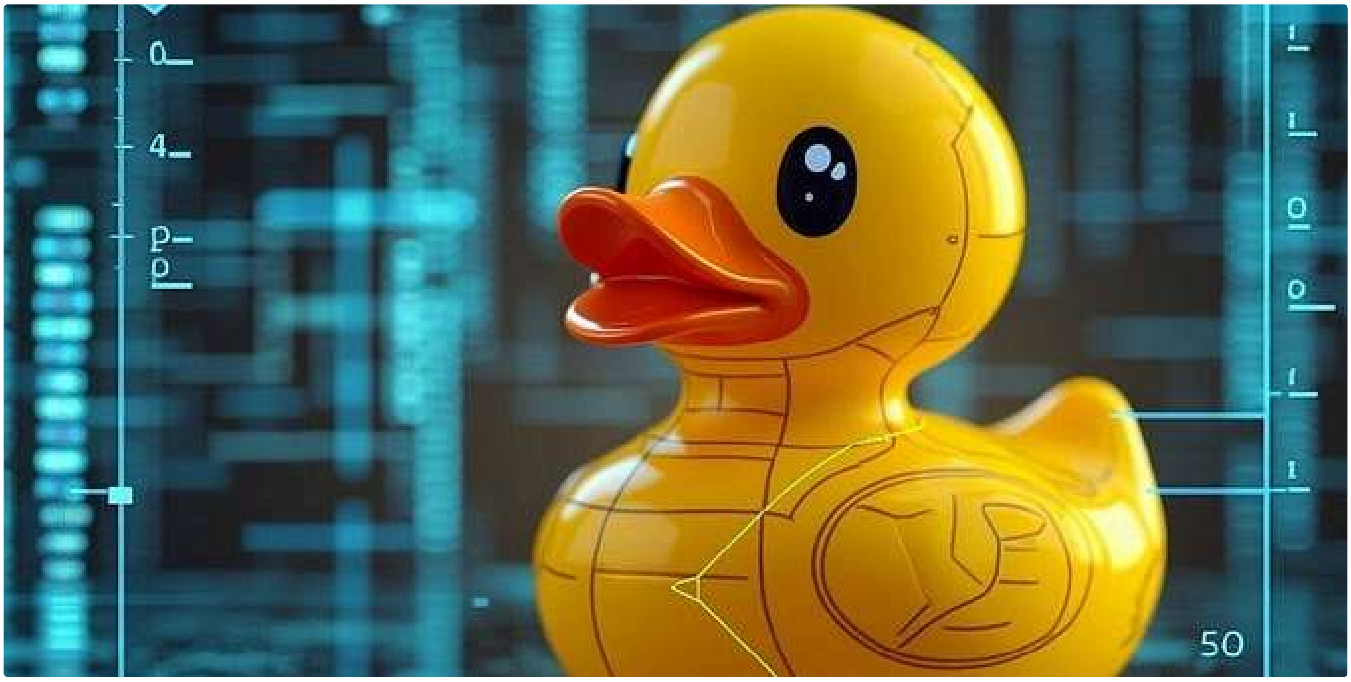
I work as PostgreSQL specialist & database reliability engineer at NetApp Deutschland, Open Source Services division.

No responses yet

What are your thoughts?

Respond

More from Josef Machytka

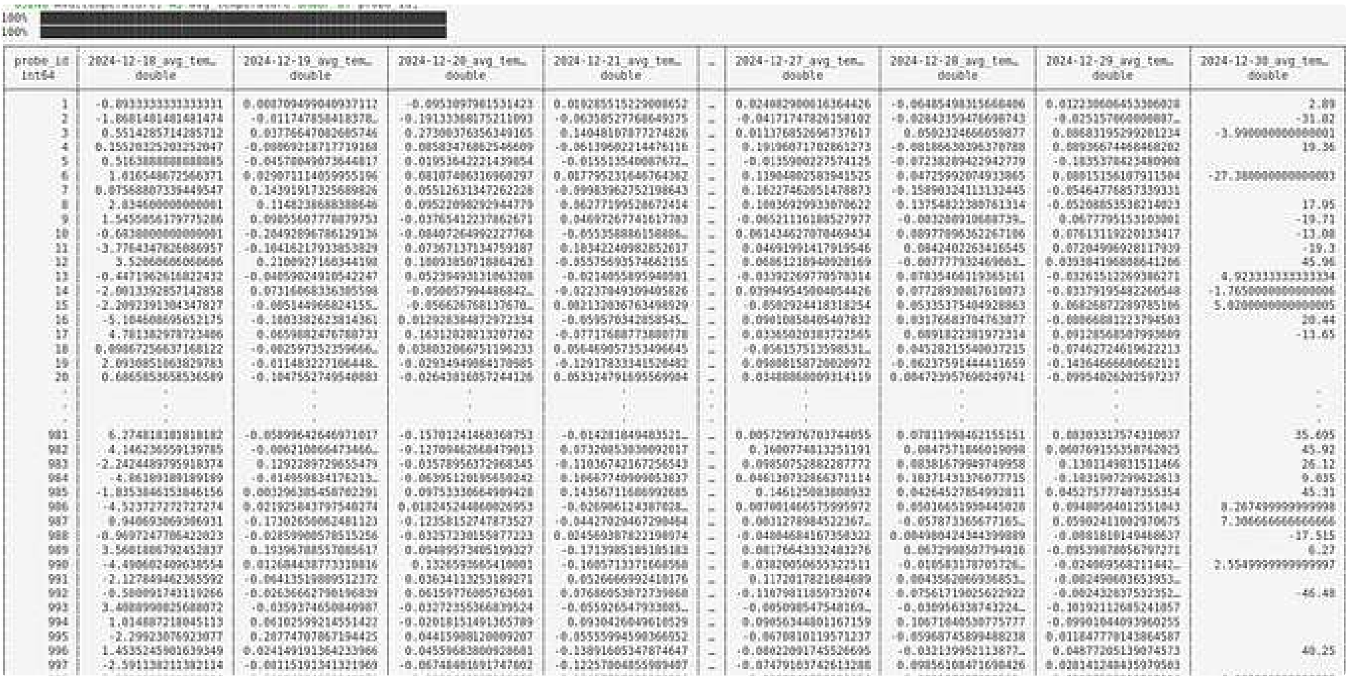


Josef Machytka

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 18 2



100%
100%

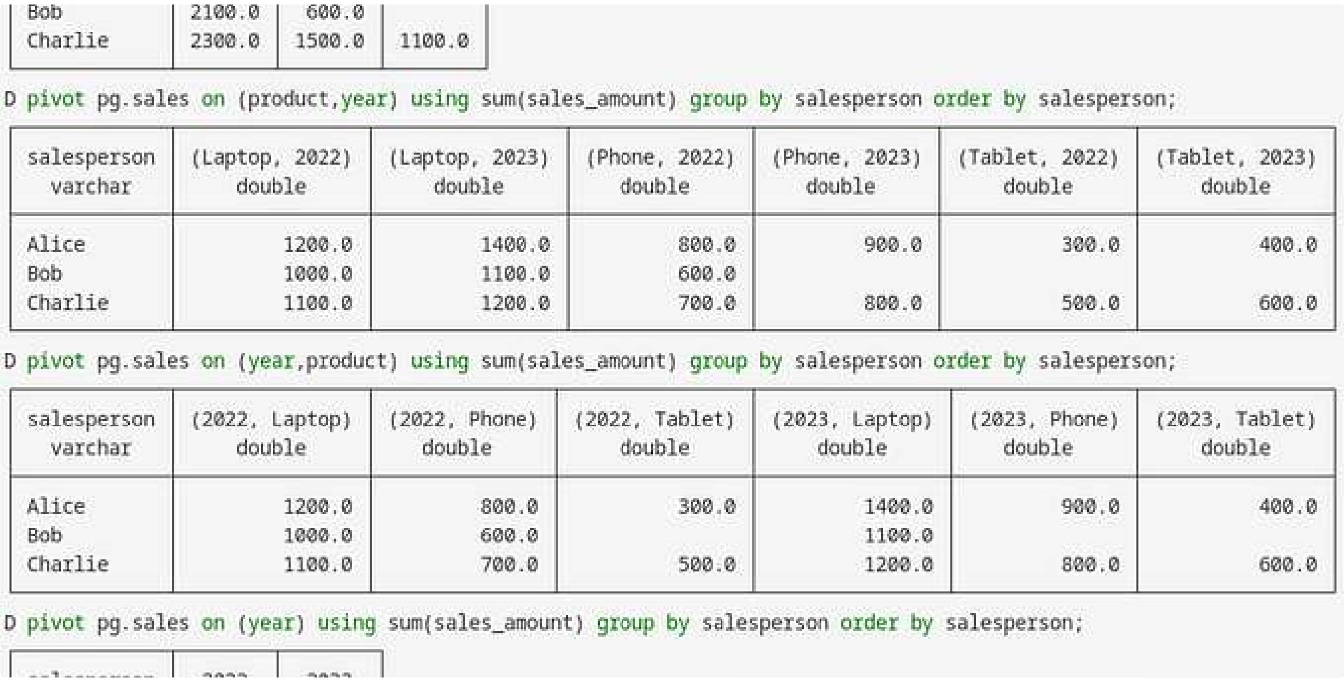
probe_id int64	2024-12-18_avg_tem… double	2024-12-19_avg_tem… double	2024-12-20_avg_tem… double	2024-12-21_avg_tem… double	…	2024-12-27_avg_tem… double	2024-12-28_avg_tem… double	2024-12-29_avg_tem… double	2024-12-30_avg_tem… double
1	-0.8933333333333331	0.008709499040937112	-0.0951097981531423	0.010285515229008652	…	0.024082900016364426	-0.06485498315668406	0.012230606653306028	2.09
2	-1.8681481481481474	-0.011747858418378…	-0.19133368175211093	-0.06358527768649375	…	-0.04171747826158102	-0.02843359476698743	-0.025157060000087…	-31.02
3	0.5514285714285712	0.03776647082685746	0.27300376356349165	0.14048107877274826	…	0.011376852696737617	0.0502324666059877	0.08683195299201234	-3.9900000000000001
4	0.15520325203252047	-0.008692187177719168	0.08583476862546609	-0.06139602214476116	…	0.19196071702861273	-0.08186630396370788	0.08936674468468202	19.36
5	0.5163888888888885	-0.04578049073644817	0.01953642221439854	-0.015513540087672…	…	-0.0135900227574125	-0.07238289422942779	-0.1835378423480908	
6	1.016548672566371	0.029071114059955196	0.08107406316960297	0.017795231646764362	…	0.11904802583941525	0.04725992074913865	0.08015156107911504	-27.380000000000003
7	0.07568807339449547	0.14391917325689826	0.05512631347262228	-0.09983962752198643	…	0.16227462051478873	-0.15890324113132445	-0.05464776857339331	
8	2.034600000000001	0.11482386883880646	0.09522098292944779	0.06277199520672414	…	0.10036929933070622	0.13754822380761314	-0.05208853538214023	17.95
9	1.5455056179775286	0.09855607770879753	-0.03765412237862671	0.04697267743617703	…	-0.06521116188527977	-0.0032089106088739…	0.0677795153103001	-19.71
10	-0.6838000000000001	-0.20492896786129136	-0.08407264992227768	-0.055358886158086…	…	0.061434627070469434	0.08977096362267106	0.07613119220133417	-13.08
11	-3.7764347826086957	-0.10416217933853829	0.07367137134759187	0.10342240982852617	…	0.04691991417919546	0.0842402263416545	0.07204996928117939	-19.3
12	3.5200000000000006	0.2100927160344198	0.10093850718864263	-0.05575693574662155	…	0.06861218940920169	-0.007777932469061…	0.039384196808641206	45.96
13	-0.4471962616822432	-0.04059024910542247	0.05239493131063208	-0.0214055895940501	…	-0.03392269770570314	0.07835468119365161	-0.03261512269386271	4.923333333333334
14	-2.0013392857142858	0.07316068336385598	-0.050057994486842…	-0.02237049309405826	…	0.039949545004054426	0.07728930817610073	-0.03379195482260548	-1.7650000000000006
15	-2.2092391304347827	-0.005144966824155…	-0.056626768137670…	0.0021132036763498929	…	-0.0502924418318254	0.05335375404928863	0.06826872289785106	5.0200000000000005
16	-5.104608695652175	-0.1803382623814361	0.012928384872972134	-0.059570342858545…	…	0.09010858405407832	0.03176683704763877	-0.008066881223794503	20.44
17	4.781382978723486	0.0659882476780733	0.16312028213207262	-0.07717688773800778	…	0.03365020383722565	0.0891822381972314	0.09128568507993609	-13.65
18	0.09867256637168122	-0.002597352359666…	0.038032066751196233	0.056469057353496645	…	-0.056157513598531…	0.04528215540037215	-0.07462724619622213	
19	2.0930851063829783	-0.011483227106448…	-0.02934949084170585	-0.12917833341520482	…	0.09808158720020972	-0.06237591444411659	-0.14364666660662121	
20	0.6865853658536589	-0.1047552749540883	-0.02643816057244126	0.053324791895569904	…	0.03488860009314119	0.004723957696249741	-0.09954026202597237	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
981	6.274818181818182	-0.05899642646971017	-0.15701241460368751	-0.014281849483521…	…	0.005729976703744055	0.07811998462155151	0.008303317574310837	35.695
982	4.146236559139785	-0.0006210066473466…	-0.12709462668479013	0.07320053030092017	…	0.16007748813251191	0.0847571846019098	0.0060769155358762025	45.92
983	-2.2424489795918374	0.1292289729655479	-0.03578956372968345	-0.11036742167256543	…	0.09850752882287772	0.08381679949749958	0.1301149031511466	26.12
984	-4.861891891891891	-0.014959834176213…	-0.06395120195650242	0.10667740909053837	…	0.046130732866371114	0.10371431376077715	-0.1031907299622613	9.035
985	-1.8353846153846156	0.003296385458702291	0.09753330664909428	0.14356711608992685	…	0.146125003808932	0.04264527854992811	0.045275777407355354	45.31
986	-4.523727272727274	0.021925843797540274	0.018245244860826953	-0.026906124387028…	…	0.007001466575995972	0.05016651959445028	0.09400504012551043	8.267499999999998
987	0.9406930693069311	-0.17302650862481123	-0.12358152747873527	-0.04427029467290484	…	0.0031278984522367…	-0.057873365677165…	0.05902411002970675	7.306666666666666
988	-0.9697247706422023	-0.02859900578515256	-0.03257230155877223	0.024560387822190974	…	-0.04804684167350322	0.004980424344399889	-0.00881810149468637	-17.515
989	3.5601886792452837	0.19396788557085617	0.09489573405199327	-0.1713985185185183	…	0.00176643332483276	0.0672990507794916	-0.09539878856797271	6.27
990	-4.490602409638554	0.012684438773310816	0.1326593665410001	-0.1605713371668568	…	0.03820050655322511	-0.010583178703726…	-0.024069568211442…	2.5549999999999997
991	-2.127849462365592	-0.006413519889512372	0.03634113253189271	0.0526666992410176	…	0.1172017821684689	0.0043562066936853…	-0.0024906036653953…	
992	-0.580091743119266	-0.02636662790196839	0.06159776085763601	-0.07686053872739868	…	-0.11079811859732074	-0.07561719025622922	-0.002432837532352…	-46.48
993	3.4088990825688072	-0.0359374650840987	-0.03272355366839524	-0.055926547933085…	…	-0.005098547548169…	-0.038956338743224…	-0.10192112685241057	
994	1.014887218045113	0.06102599214551422	-0.02018151491365789	0.09304260490610529	…	0.09056344801167159	0.10671040530775777	-0.09901044093960255	
995	-2.29923076923077	0.20774707867194425	0.04415908120009207	-0.05555994590366952	…	-0.06708101219571237	-0.05968745899488238	0.011847770143864587	
996	1.4535245901639349	0.024149191364233966	0.04559683800928681	-0.13891605347874647	…	-0.08022091745526695	-0.032139952113877…	0.04877205139074573	40.25
997	-2.591138211382114	-0.00115191341321969	-0.06748481691747602	-0.12257004855989407	…	-0.07479163742613288	0.09856108471698426	0.020141248435979503	

Josef Machytka

DuckDB Performance Problems with Inappropriate Pivoting Queries on Very Large Datasets

The DuckDB documentation clearly states that this tool is designed for handling datasets fitting into memory. I fully understand that I'm...

Jan 3 👏 6



```
│ 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 │ (Laptop, 2022) │ (Laptop, 2023) │ (Phone, 2022) │ (Phone, 2023) │ (Tablet, 2022) │ (Tablet, 2023) │
│   varchar   │    double     │    double     │    double    │    double    │    double     │    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 │ (2022, Laptop) │ (2022, Phone) │ (2022, Tablet) │ (2023, Laptop) │ (2023, Phone) │ (2023, Tablet) │
│   varchar   │    double     │    double    │    double     │    double     │    double    │    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 👏 1

Josef Machytka

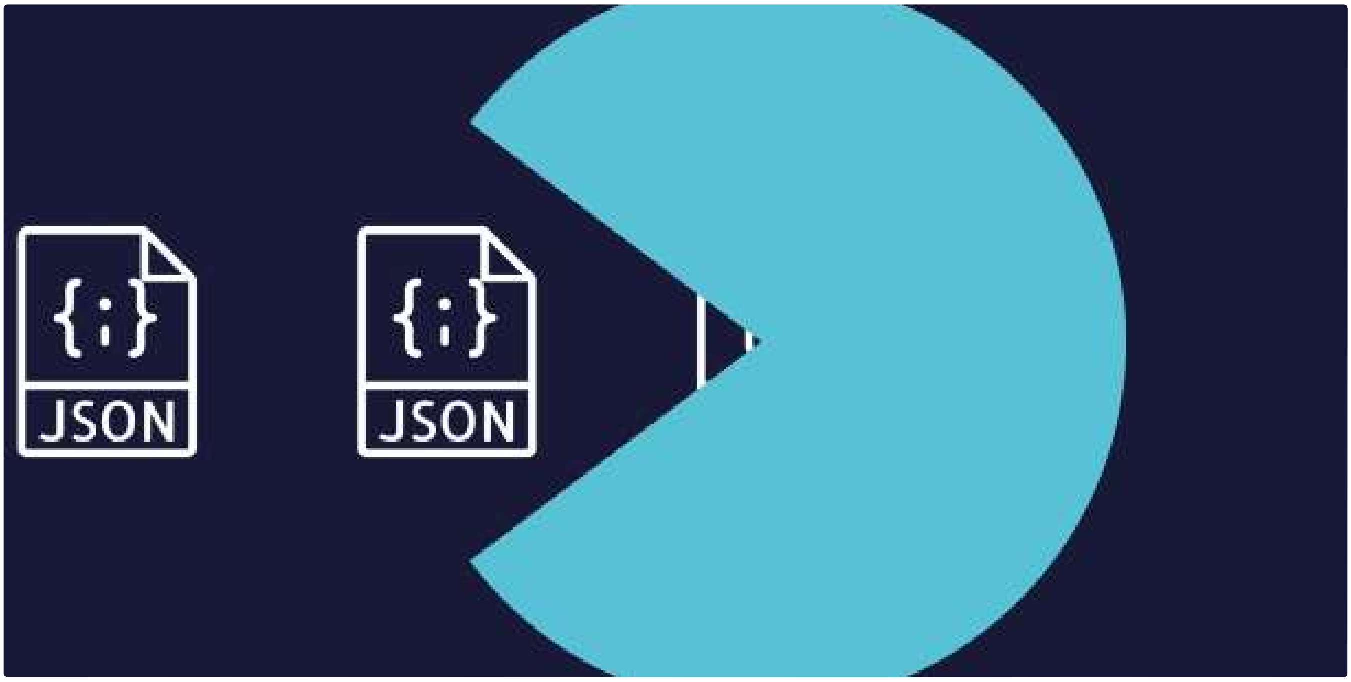
PostgreSQL and DuckDB: Supercharging Ad-Hoc Data Analysis and ETL

On Wednesday, January 8th, 2025, I had the amazing opportunity to present my talk about the DuckDB database online at the “Postgres MeetUp...

Jan 9 👏 14 💬 1

See all from Josef Machytka

Recommended from Medium

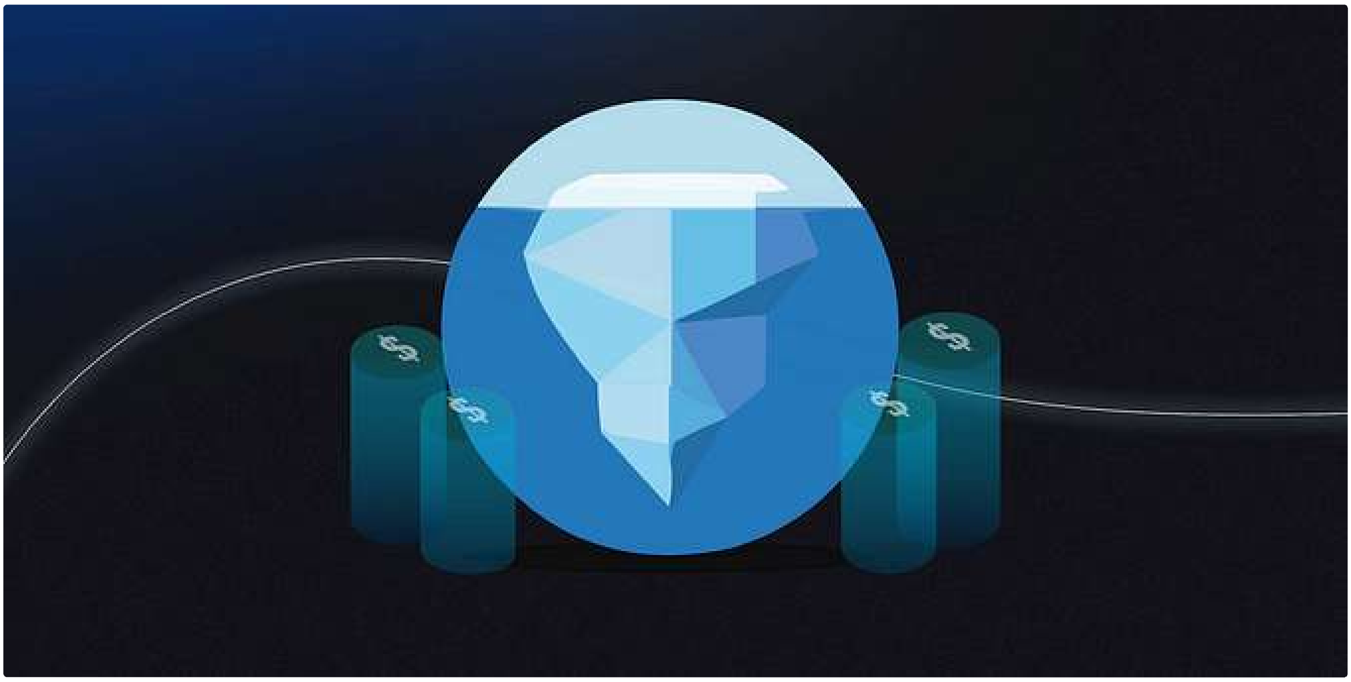


Jordi Puig

DLT is Great!

Learn the basics of dlt with this simple tutorial and forget about the data engineering pains of extracting and loading data.

Oct 12, 2024 👏 85 💬 2



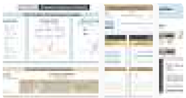
In Towards Dev by RisingWave Labs

The True Power of Apache Iceberg: Revolutionizing Modern Data Architectures

Apache Iceberg is more than just a new technology; it's a paradigm shift in how we manage and utilize data.

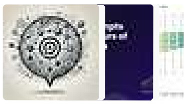
4d ago

Lists



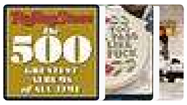
Practical Guides to Machine Learning

10 stories · 2182 saves



ChatGPT prompts

51 stories · 2522 saves



Staff picks

806 stories · 1601 saves

Natural Language Processing

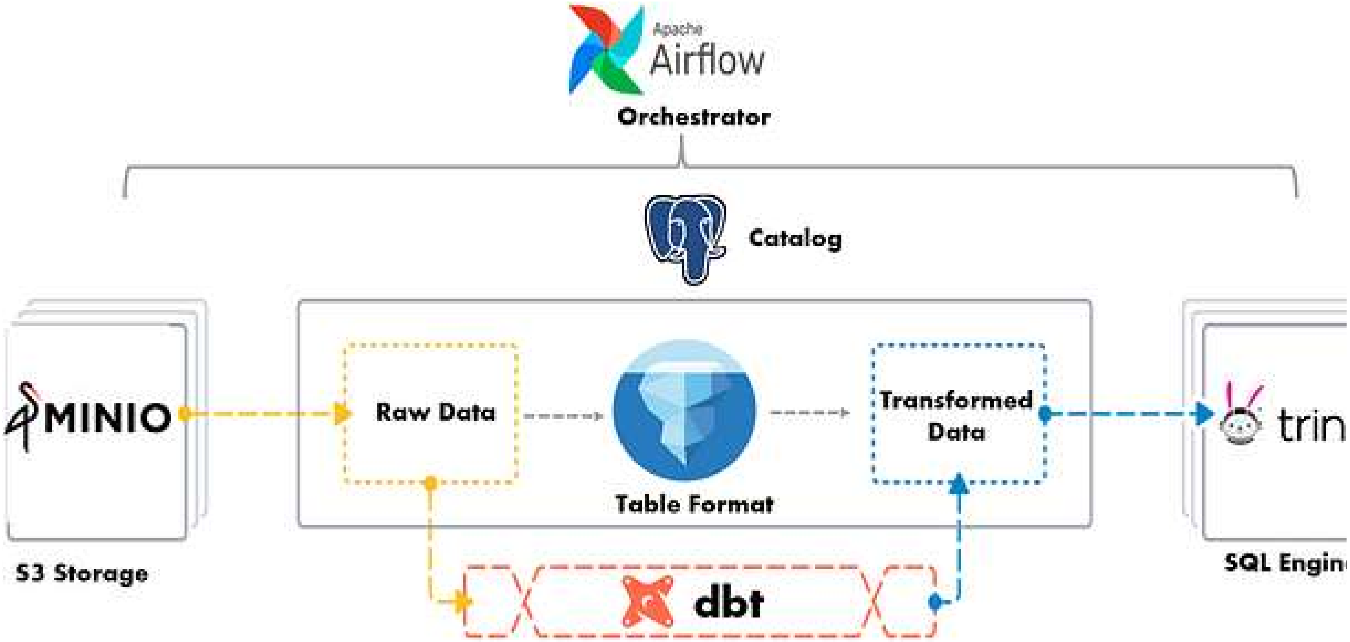
1908 stories · 1563 saves

In Level Up Coding by Anna Geller

2025 Data Engineering & AI Trends

How GenAI, new data regulations, Postgres, DuckDB, and open table formats affect data engineering in 2025 and beyond

Jan 24 216 2



In Dev Genius by Haq Nawaz

Building an End-to-End Data Lake ELT Pipeline using Modern Data Stack

Using Airflow, dbt, Iceberg, MinIO, and Trino

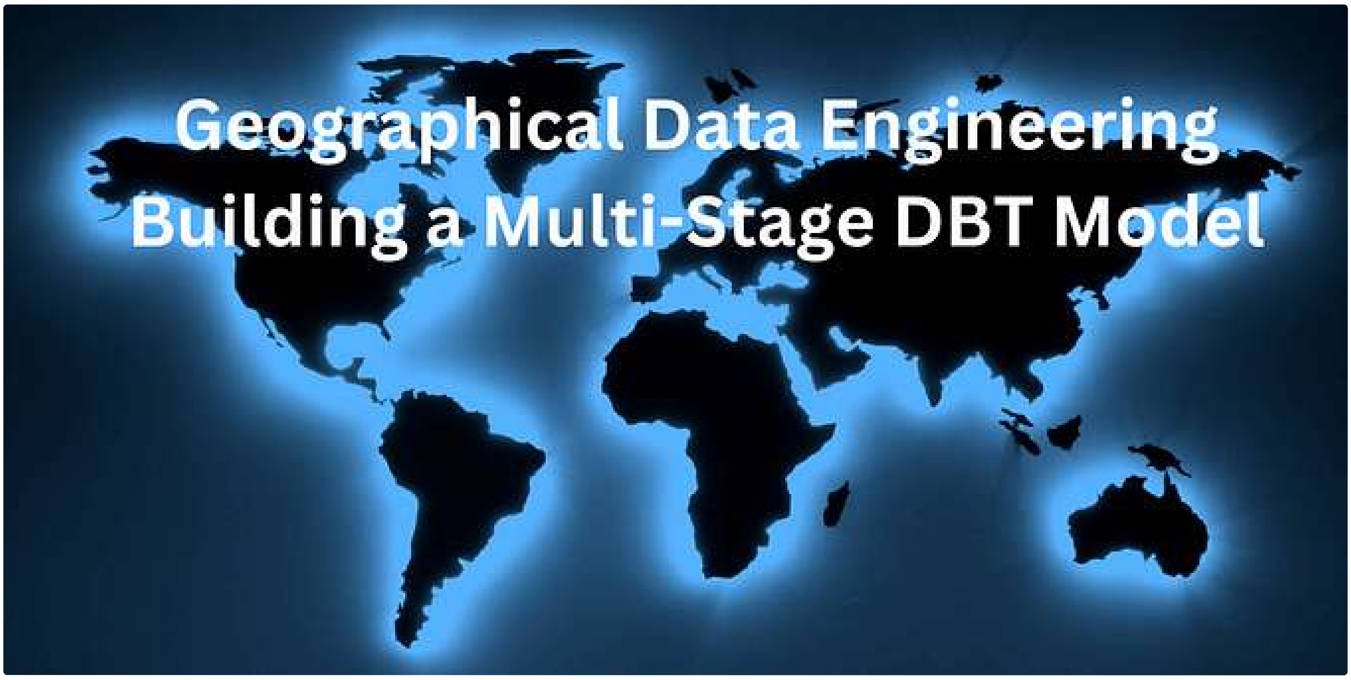
4d ago 117

Nkem Onyemachi

Data Load Tool (dlt) Series 2: Creating Pipelines from Different Sources

In the first article of this series,

6d ago 5


Geographical Data Engineering
Building a Multi-Stage DBT Model

Data Dev Backyard

Geographical Data Engineering: Building a Multi-Stage DBT Model and Querying Using DuckDB

In the previous tutorials, we have introduced DBT, and discussed the core concepts behind DBT. In this tutorial, we would like to create...

5d ago 1

See more recommendations