



DATA SATURDAY

BELO HORIZONTE

23 de Novembro de 2024

08:00 às 16:30

PUC Minas - Coração Eucarístico

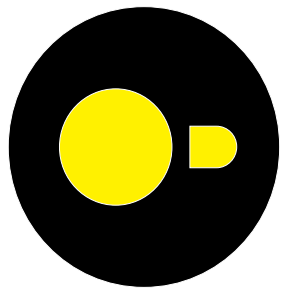
Rua Dom José Gaspar, 500, Prédio 43, Coração Eucarístico
Belo Horizonte, MG





DATA SATURDAY

BELO HORIZONTE



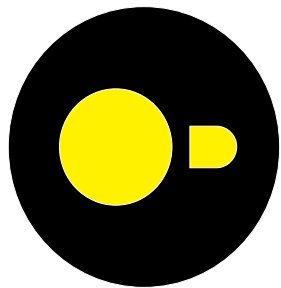
DuckDB

O Patinho no Lago de Dados



DATA SATURDAY

BELO HORIZONTE



DuckDB

O Patinho **IA** no Lago de Dados



DATA SATURDAY

BELO HORIZONTE



Patrocinadores



Danilo Santos

Graduado em Sistemas de Informação, especialização em Engenharia de Dados.



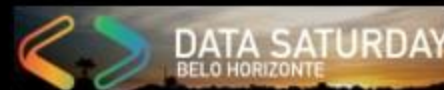
Supervisor de Dados.



Professor de Banco de Dados e Python.



Projeto de consultoria, freelancer e treinamento.





Agenda?

- 🦆 O que é DuckDB?
- 🦆 Principais Características
- 🦆 Documentação
- 🦆 Livro MutherDuck
- 🦆 Hands on



O que é DuckDB?

- 🐥 CWI Amsterdam, Holanda, 2018
- 🐥 É um banco de dados relacional
- 🐥 Processamento Analítico - OLAP
- 🐥 Embarcado
- 🐥 Processamento em tempo de execução
- 🐥 Código Aberto
- 🐥 Gratuito



Porque DuckDB?



Hannes Mühleisen



Porque DuckDB?

Os patos são muito versáteis



Voam



Andam



Nadam



Principais Características



Simple de usar



In-Memory



Simple de instalar



Vetorização



Sem Servidor



Paralelo



Arquivo Único



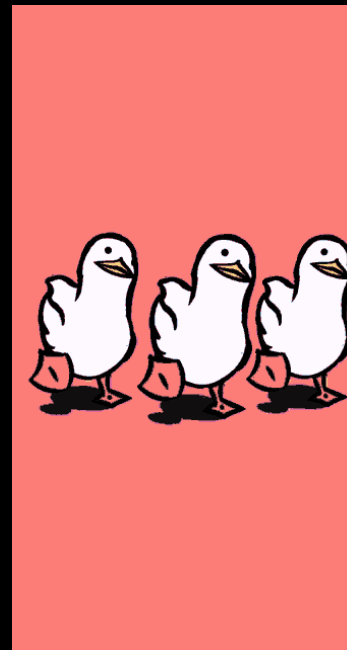
Nunca Falha *



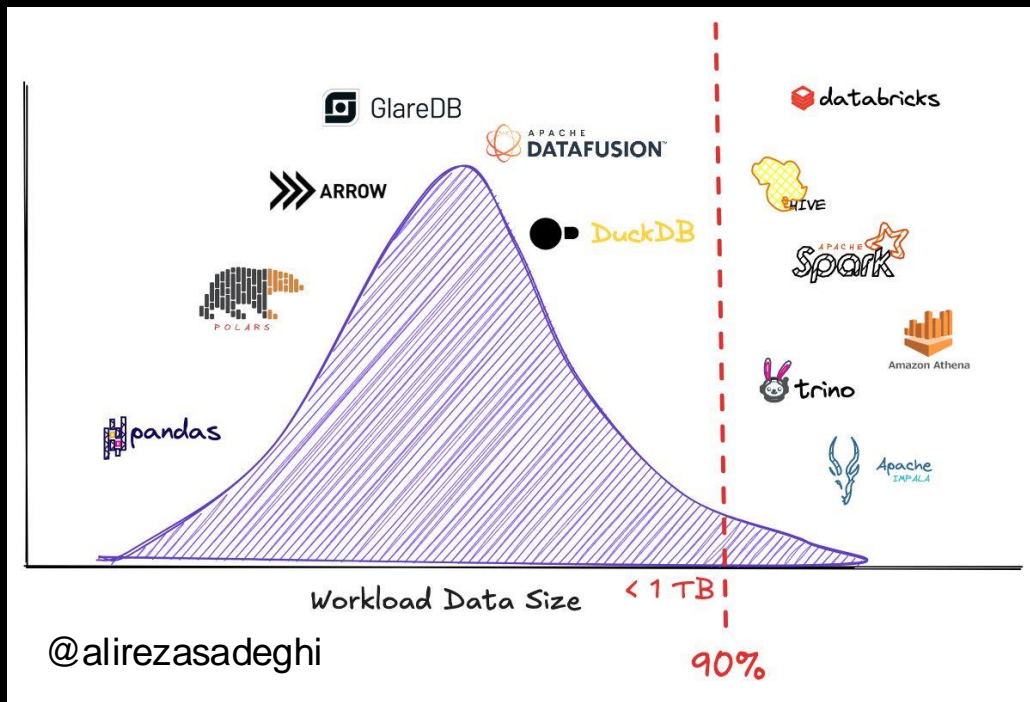
GBs



TBs



Principais Características



Principais Características

Otimização de Query



Reescrita de execução



Join, Agregação, Ordenação



Achatamento de subquery



Projeção e filtros



Principais Características

SGBDS

OLTP

Processamento em linha

SQL

Sem Data Frame

Servidor

Armazenamento complexo

Pandas

OLAP

Processamento em Coluna

Sem SQL

Data Frame

Sem servidor

Sem armazenamento

DuckDB

OLAP

Processamento Vetorizado

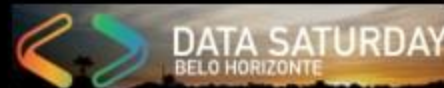
SQL

Data Frame

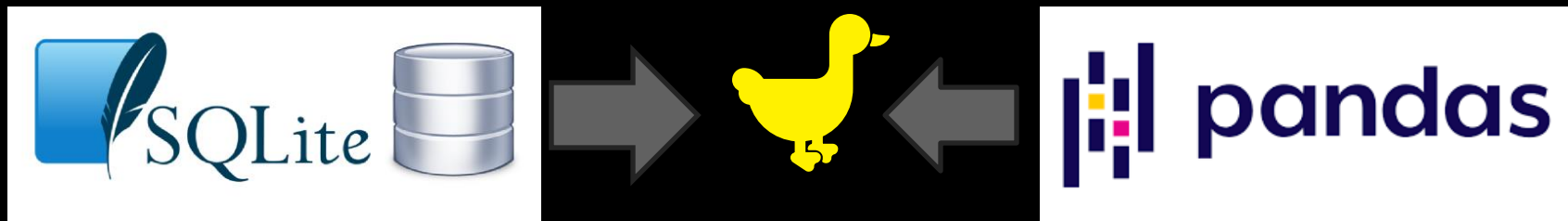
Sem Servidor

Armazena único arquivo

Otimizado cache CPU



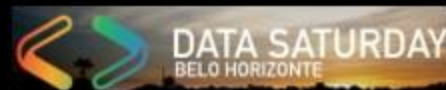
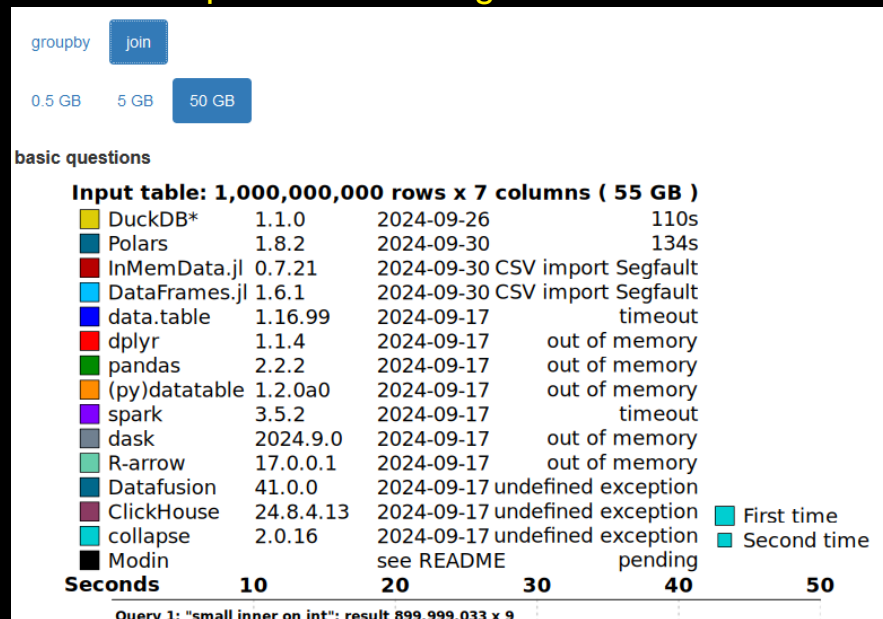
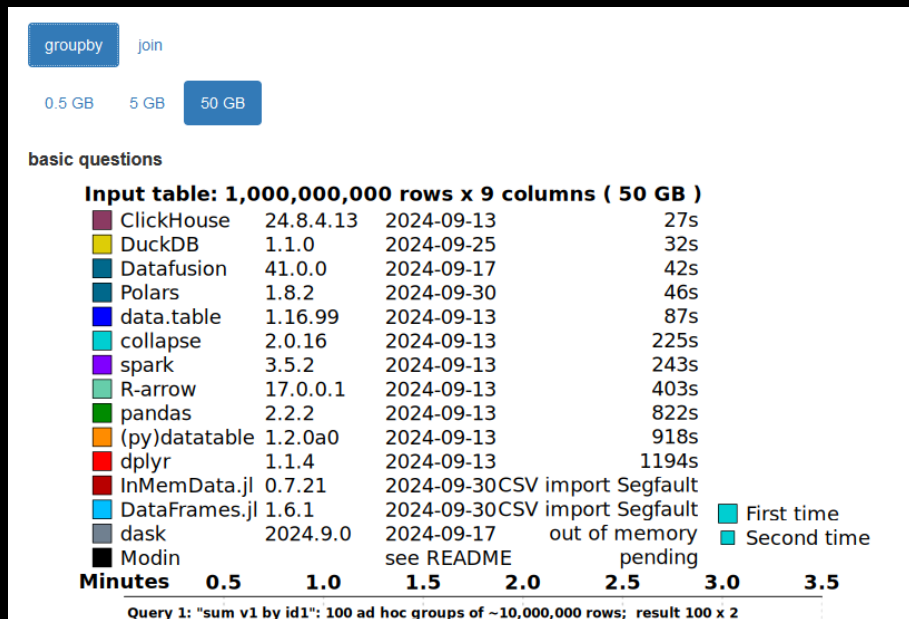
Principais Características



Principais Características



<https://duckdblabs.github.io/db-benchmark/>



Principais Características

Python DB API

 Engine PostgreSQL

API Relacional

 **NumPy**

 **PyArrow**

 **Pandas**

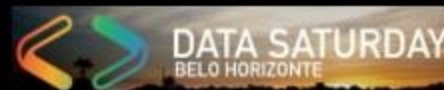
 **Polars**

 **SQLAlchemy**

**** zero cópia**



Principais Características



Principais Características



Principais Características



Principais Características

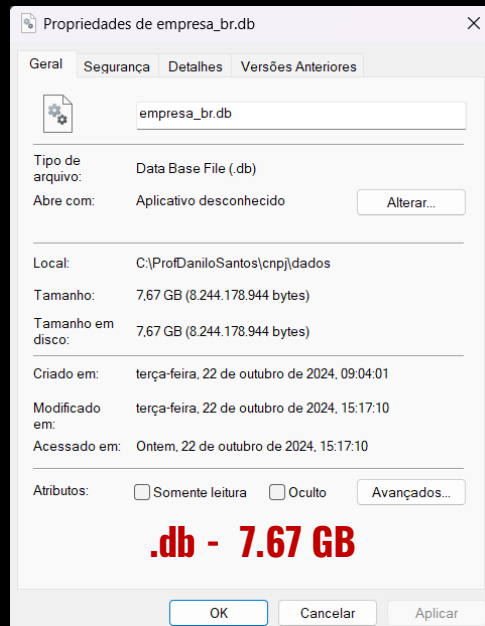
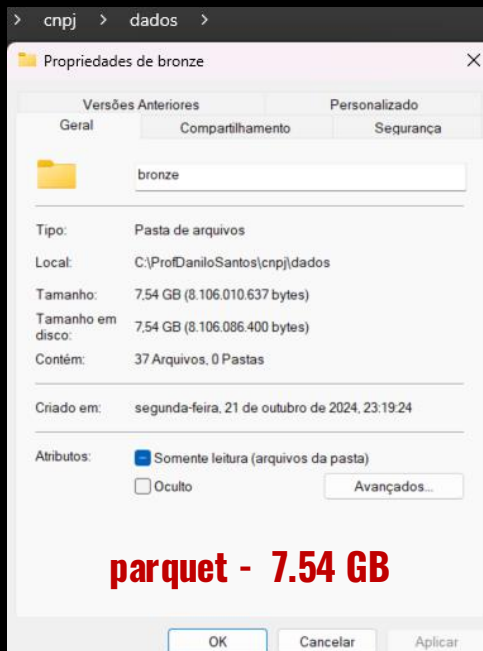
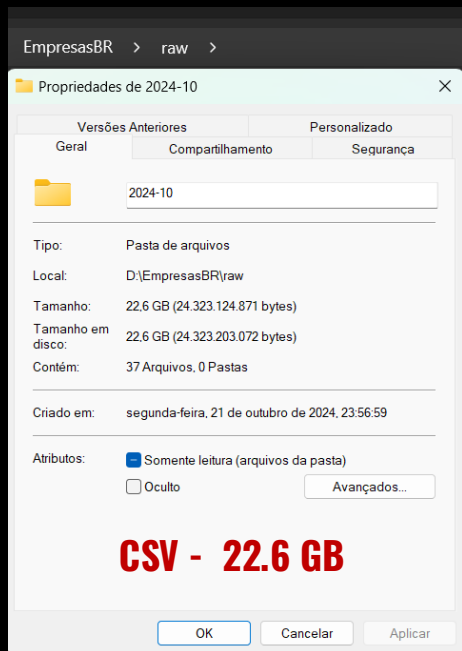


Compressão de Dados

Version	Taxi	On Time	Lineitem	Notes	Date
DuckDB v0.2.8	15.3GB	1.73GB	0.85GB	Uncompressed	July 2021
DuckDB v0.2.9	11.2GB	1.25GB	0.79GB	RLE + Constant	September 2021
DuckDB v0.3.2	10.8GB	0.98GB	0.56GB	Bitpacking	February 2022
DuckDB v0.3.3	6.9GB	0.23GB	0.32GB	Dictionary	April 2022
DuckDB v0.5.0	6.6GB	0.21GB	0.29GB	FOR	September 2022
DuckDB dev	4.8GB	0.21GB	0.17GB	FSST + Chimp	<code>now()</code>
CSV	17.0GB	1.11GB	0.72GB		
Parquet (Uncompressed)	4.5GB	0.12GB	0.31GB		
Parquet (Snappy)	3.2GB	0.11GB	0.18GB		
Parquet (ZSTD)	2.6GB	0.08GB	0.15GB		

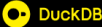


Compressão de Dados





Documentação

 DuckDB

Q Search

ctrl+k

Installation

Documentation

Getting Started

Connect

Overview

Concurrency

Data Import

Overview

Data Sources

CSV Files

JSON Files

Multiple Files

Parquet Files

Partitioning

Appender

INSERT Statements

Client APIs

SQL

Configuration

Extensions

Guides

Operations Manual

Development

Internals

Sitemap

What's DuckDB

DuckDB Installation

Light Mode 1.1 (stable)

This page contains installation options for DuckDB. For production use, we recommend the stable release.

Binaries are available for major programming languages and [platforms](#). If there are no pre-packaged binaries available, consider [building DuckDB from source](#).

Version

Stable release (v1.1.2)

Nightly build (bleeding edge)

Environment

Command line

Python

R

Java

Node.js

Rust

Go

C/C++

ODBC

Platform

Windows

macOS

Linux

Download method

Package manager

Direct download

Architecture

x86_64

arm64

Installation

winget install DuckDB.cli

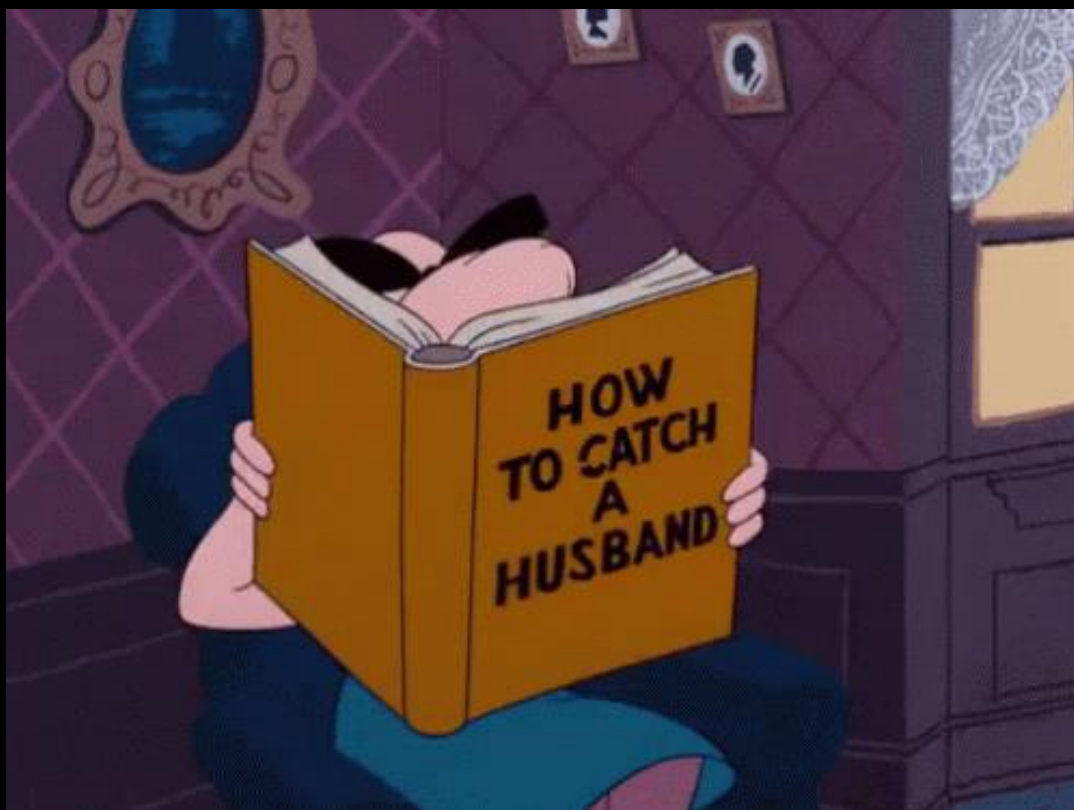
Note: Each DuckDB client is installed without relying on any other DuckDB clients. For example, the Python library can use a different version than the CLI client. Therefore, they need to be updated separately.

Usage example

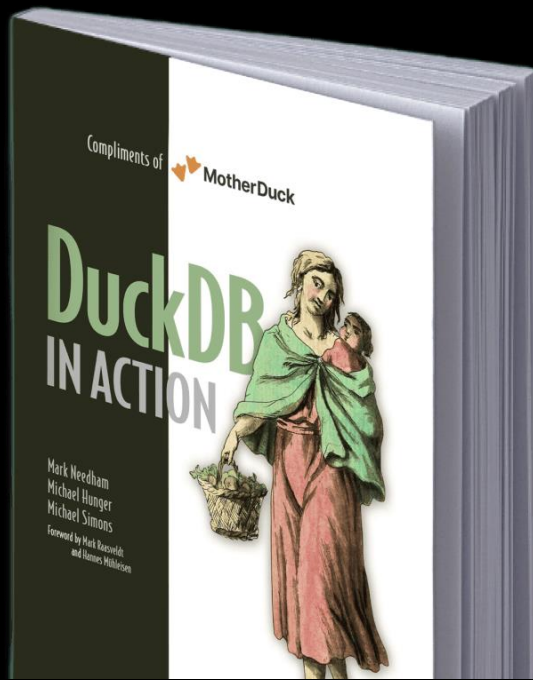
duckdb

<https://duckdb.org/>





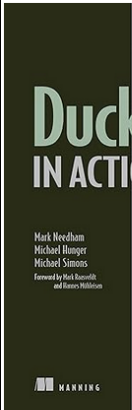
DuckDB In Action - Mother Duck



<http://motherduck.com/duckdb-book-brief>



DuckDB In Action - Mother Duck



Mark Needham
Michael Hunger
Michael Simons
Foreword by Mark Needham
and Eamon Healy

MANNING

Duckdb in Action

Capa comum – 27 agosto 2024

Edição Inglês | por [Mark Needham](#) (Autor), [Michael Hunger](#) (Autor), [Michael Simons](#) (Autor)

[Ver todos os formatos e edições](#)

Em até 12x de R\$ 52,91 com juros [Ver parcelas disponíveis](#) ▼


Dive into DuckDB and start processing gigabytes of data with ease--all with no data warehouse.






DuckDB is a cutting-edge SQL database that makes it incredibly easy to analyze big data sets right from your laptop. In *DuckDB in Action* you'll learn everything you need to know to get the most out of this awesome tool, keep your data secure on prem, and save you hundreds on your cloud bill. From data ingestion to advanced data pipelines, you'll learn everything you need to get the most out of DuckDB--all through hands-on examples.

Open up *DuckDB in Action* and learn how to:

- Read and process data from CSV, JSON and Parquet sources both locally and remote
- Write analytical SQL queries, including aggregations, common table expressions, window functions, special types of joins, and pivot tables
- Use DuckDB from Python, both with SQL and its "Relational"-API, interacting with databases but also data frames
- Prepare, ingest and query large datasets
- Build cloud data pipelines


▼ [Leia mais](#)

 [Relatar um problema com este produto](#)

Número de páginas	Idioma	Editora	Data da publicação	Dimensões
 312 páginas	 Inglês	 Manning Publications	 27 agosto 2024	 18.73 x 1.78 x 23.5 cm

[Passe o mouse para ampliar a imagem](#)

[Ler amostra](#)



Kindle
R\$ 241,64

Capa Comum
R\$ 566,90

[Outros Usado e Novo a partir de R\\$ 562,98](#) ▼

R\$ 566⁹⁰

Entrega GRÁTIS: 10 - 27 de Dezembro. [Ver detalhes](#)

📍 [Enviar para Danilo - Juiz De Fora 36036230](#)

Em estoque

Compra Internacional

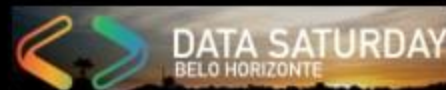
Os [tributos de importação](#) estão incluídos. Você não terá custos extras.

[Adicionar ao carrinho](#)

[Comprar agora](#)

Enviado por GF Books, Inc.
Vendido por GF Books, Inc.
Devolução Elegível para Devolução ou Reembolso em até 7 dias após o recebimento

Pagamento [Transação segura](#)





Falando?

DUCKDB: UM SGBD EMBARCADO PARA ANÁLISE/CIÊNCIA DE DADOS
Insight Lab • 1,3 mil visualizações • Transmitido há 3 anos

Omega Talks | Pedro Holanda • DuckDB: Um SGBD embarcado para análise/ciência de dados
Ômega Data Science • 713 visualizações • Transmitido há 3 anos

Processamento de Dados com DuckDB, AWS e Spark com Pedro Holanda
Iury Rosal • 700 visualizações • Transmitido há 5 meses

Instalando Configurando DuckDB DBEaver
Carreira Dados • 241 visualizações • há 5 meses

Supletivo DH - DuckDB: O Banco de Dados Analítico de Alta Velocidade
Rodrigo TEORIA • 4 mil visualizações • Transmitido há 1 ano

Como utilizar o DuckDB com Python: De Excel a Parquet com DuckDB!
DataWay BR • 3,3 mil visualizações • há 8 meses

DuckDB: Introdução (Gua Definitivo e Completo)
Iury Rosal • 1,7 mil visualizações • Transmitido há 7 meses

Demetrius Mata • Seguindo
Engenheiro de Dados SR | Arquiteto de Dados Azure | AWS | Databricks
5 m •

Sua dose sobre o mundo dos dados
#DuckDB #ProcessamentoDeDados #AnaliseDeDados #BigData #Mundo dos Dados

Luciano Vasconcelos Filho • 1º
Engenheiro de dados e professor da Jornada
Acesse meu site
9 m •

Você trabalha com SQL?
Se sim, tenho certeza que ta se

Iury R. (He/his) • Seguindo
Engenheiro de Dados | Mestrado em Informática pela PUC-RIO | Python, SQL...
5 m • Editado •

...uma aula que produzi sobre DuckDB, dessa vez me dentro da TechWeek da **Avenue Code**. Essa é ...mais

Rodrigo Santana • Seguindo
Data Engineer | Liderança em Dados e Inteligência Artificial
5 m •

Aqui na empresa, implementamos o **#DuckDB** em alguns de nossos clientes resultados superaram nossas expectativas.

Na minha opinião, essas são algun

Téo Calvo • Seguindo
Eu transformo vidas por meio do ensino
2 m •

DUCKDB INTEGRADO NO SEU NOTEBOOK!!
Você já usou ou já ouviu falar do DuckDB? Vem comigo, que nesse post eu ...mais

Daniel S. • 2º
Data Engineer | Analytics Engineer | SQL | Python | Spark | Airf...
10 m •

Vale a pena dar uma olhada! DuckDB já é mega performático, com tu
Só alegria!

Anderson Amaral
Building Client-Tested
4 d •

Ainda não viu ninguém usan

DuckDB
29.392 seguidores
10 m •

We wrote a performance guide for DuckDB users! This guide covers

Charles Lima • 1º
Analytics Engineer | Data Engineer | Data Vault 2.0 | Databricks | Power BI | P...
3 sem • Editado •

Quando eu falo de DuckDB as pessoas pensam que é só mais um banco de dados

JOVIANO SILVEIRA • 1º
Microsoft MVP | Controller - Contador | POWER QUERY | Power BI | SQL | Py...
Acesse meu site
1 sem •

leia **INSERIR DADOS DE PLANILHAS NO BANCO DE DADOS**
Se você é daqueles que se pergunta "como armazenar meus dados sem toda aquela parafernália de servidor?", SQLite e DuckDB estão aqui para salvar ...mais

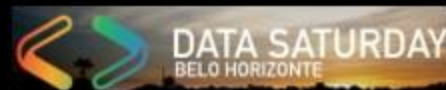


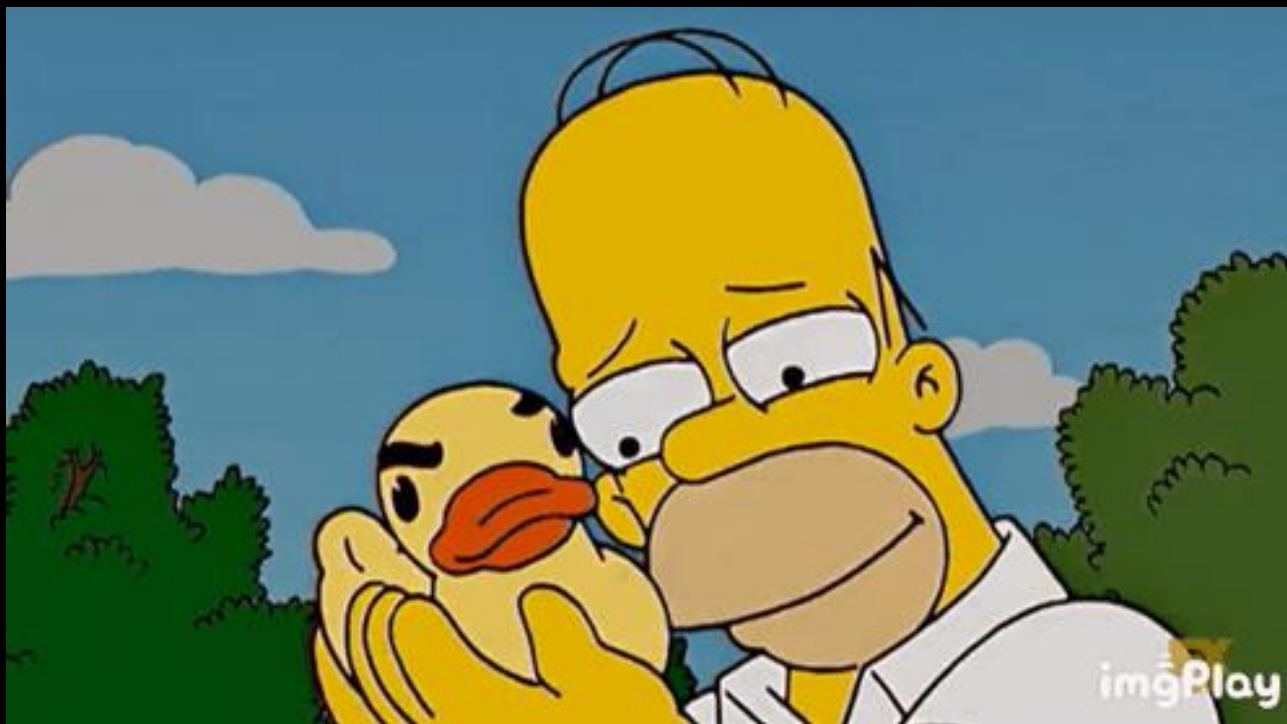
Falando?

<https://db-engines.com/en/ranking>

423 systems in ranking, October 2024

Rank			DBMS	Database Model	Score		
Oct 2024	Sep 2024	Oct 2023			Oct 2024	Sep 2024	Oct 2023
1.	1.	1.	Oracle +	Relational, Multi-model ⓘ	1309.45	+22.85	+48.03
2.	2.	2.	MySQL +	Relational, Multi-model ⓘ	1022.76	-6.73	-110.56
3.	3.	3.	Microsoft SQL Server	Relational, Multi-model ⓘ	802.09	-5.67	-94.79
4.	4.	4.	PostgreSQL +	Relational, Multi-model ⓘ	652.16	+7.80	+13.34
5.	5.	5.	MongoDB +	Document, Multi-model ⓘ	405.21	-5.02	-26.21
6.	6.	6.	Redis +	Key-value, Multi-model ⓘ	149.63	+0.20	-13.33
7.	7.	↑ 11.	Snowflake +	Relational	140.60	+6.88	+17.36
8.	8.	↓ 7.	Elasticsearch	Multi-model ⓘ	131.85	+3.06	-5.30
9.	9.	↓ 8.	IBM Db2	Relational, Multi-model ⓘ	122.77	-0.28	-12.10
10.	10.	↓ 9.	SQLite	Relational	101.91	-1.43	-23.23
57.	↑ 60.	↑ 98.	DuckDB	Relational	5.98	+0.37	+2.34





Vagas?





Caso de uso

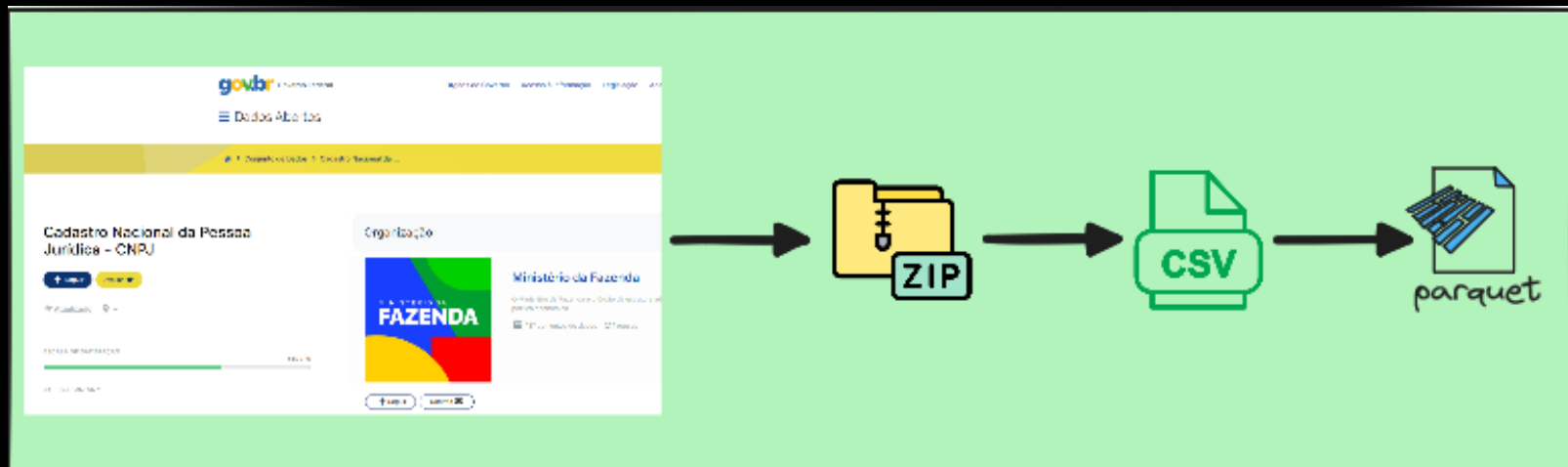
- 🐦 Análise de Cadastro Nacional de Pessoas Jurídicas, dados abertos do Gov.BR;
- 🐦 Apuração de concurso de rádio amador;
- 🐦 Integração de sistema entre APIs/Json e Mysql;
- 🐦 Documentação de pipelines Data Factory (poc);
- 🐦 Migrando o curso de BD para DuckDB;



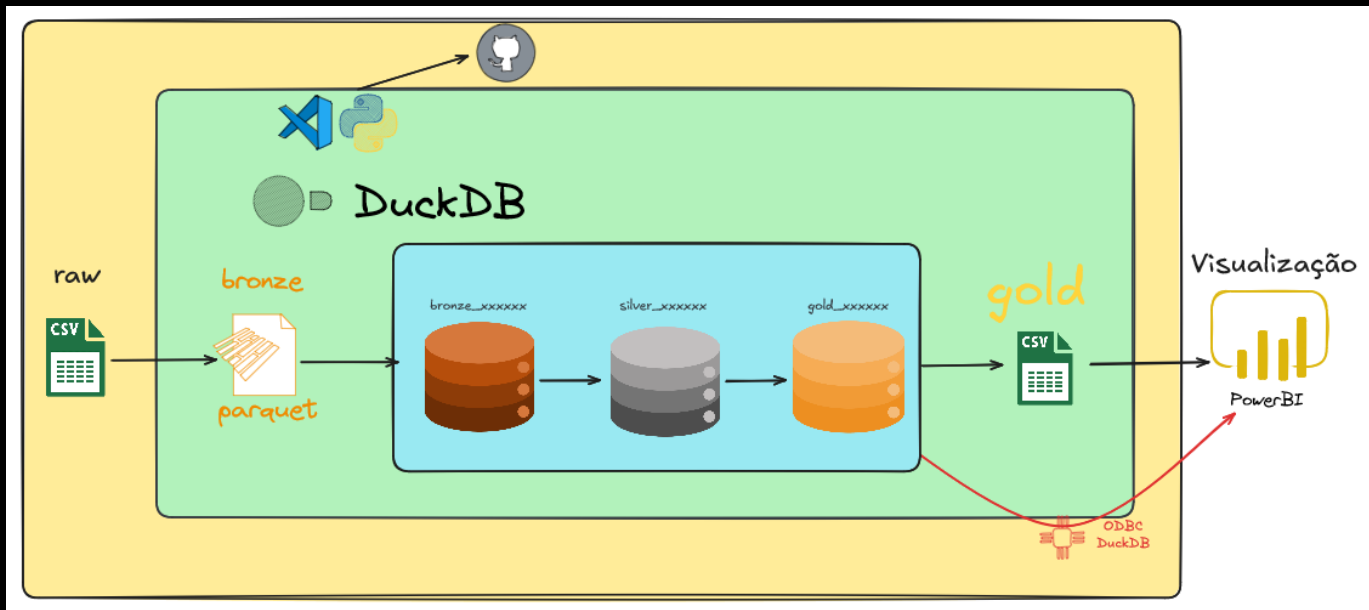


Hands on

<https://dados.gov.br/dados/conjuntos-dados/cadastro-nacional-da-pessoa-juridica---cnpj>

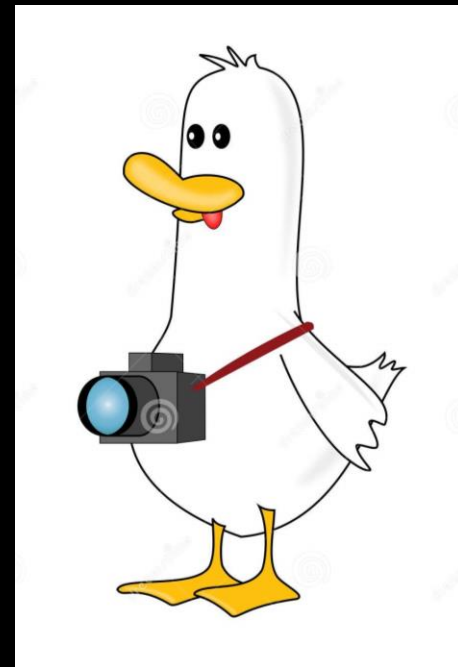


Hands on









Arquivos

https://github.com/profdanilosantos/duckdb_data_saturday_2024_belo Horizonte



Danilo Santos

- Email: danilo@danilosantos.dev.br
- linkedin.com/in/danilo-oliveira-santos
- @djkalango

