



DuckDB

O Patinho no Lago de Dados

Danilo Santos



Data Saturday
Vitória 2024

Patrocinadores

GOLD



Data Saturday
Vitória 2024

Patrocinadores

SILVER



BRONZE



Data Saturday
Vitória 2024

Danilo Santos

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

Foco de carreira em engenharia de dados, profissional da área de tecnologia com mais de 20 anos de experiência, atuei com programação, datacenter, manutenção e servidores Linux.

Vero Internet: Supervisor de Dados.

Faculdade Ensin-e: Professor de Banco de Dados e Python.

Menara Dados: meu projeto de consultoria, freelancer e treinamento.

Certificação: AZURE AZ-900



DuckDB



Data Saturday
Vitória 2024



DuckDB



Data Saturday
Vitória 2024

O que vamos falar?

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



Data Saturday
Vitória 2024

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



Data Saturday
Vitória 2024

Porque se chama DuckDB?



Hannes Mühleisen



DuckDB



Data Saturday
Vitória 2024

Porque se chama DuckDB?

Os patos são muito versáteis



Voam



Andam



Nadam



DuckDB



Data Saturday
Vitória 2024

Principais Características



Simple de usar



In-Memory



Simple de instalar



Vetorização



Sem Servidor



Paralelo



Arquivo Único



Nunca Falha *



DuckDB



Data Saturday
Vitória 2024

Principais Características

Otimização de Query



Reescrita de execução



Join, Agregação, Ordenação



Achatamento de subquery



Projeção e filtros



DuckDB



Data Saturday
Vitória 2024

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

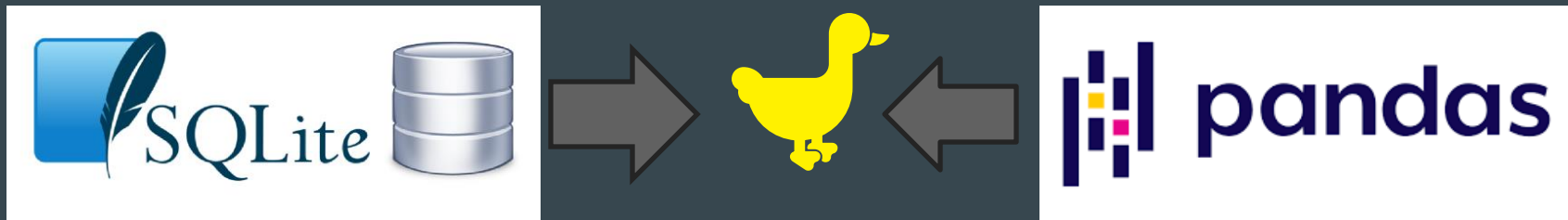
Data Frame

Sem Servidor

Armazena único arquivo

Otimizado cache CPU

Principais Características



Principais Características

API SQL

- * PostgreSQL

API Relacional

- * NumPy
- * PyArrow
- * Pandas
- * Polars
- * SQL Alchemy

** zero cópia

Principais Características



DuckDB



Data Saturday
Vitória 2024

Principais Características



DuckDB



Data Saturday
Vitória 2024

Principais Características



DuckDB



Data Saturday
Vitória 2024

Principais Características



Data Saturday
Vitória 2024

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		

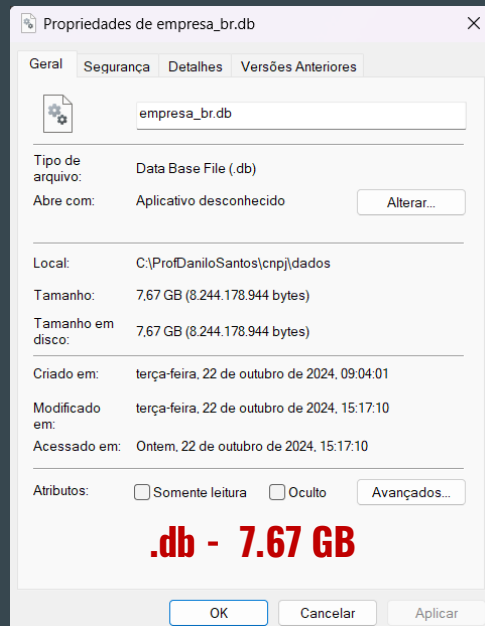
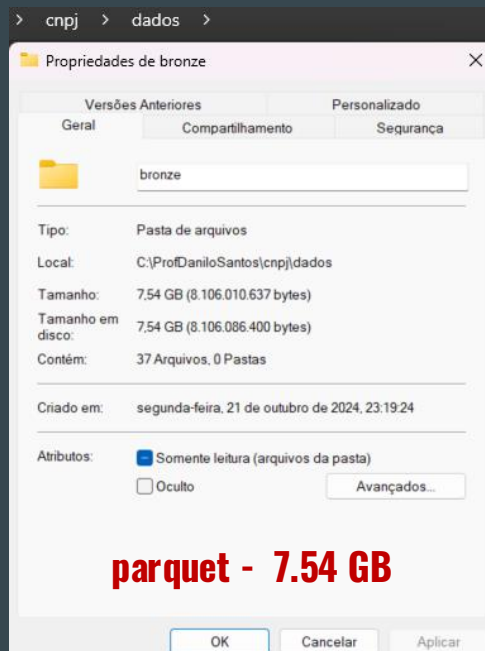
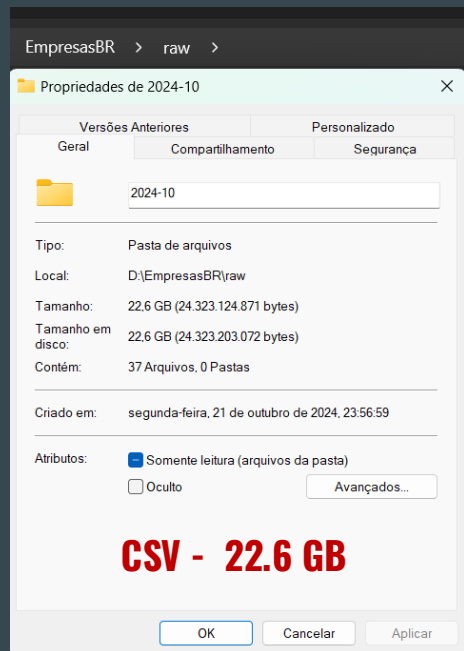


DuckDB

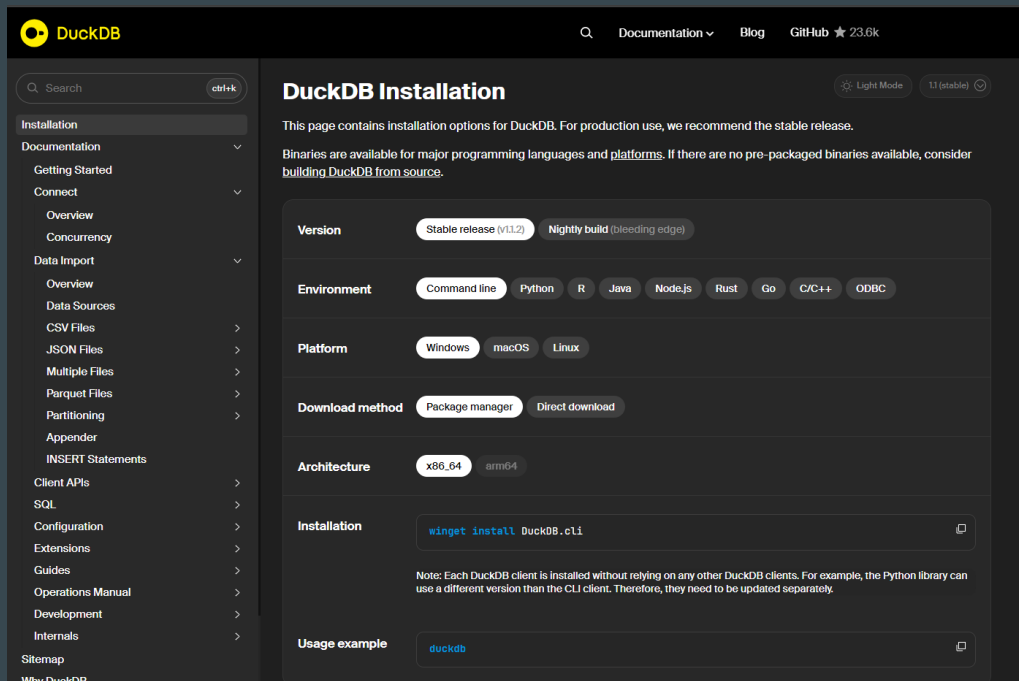


Data Saturday
Vitória 2024

Compressão de Dados



Documentação



The screenshot shows the DuckDB documentation website. The top navigation bar includes the DuckDB logo, a search icon, and links for Documentation, Blog, and GitHub (23.6k stars). The left sidebar contains a menu with categories like Installation, Documentation, Connect, Data Import, Client APIs, SQL, Configuration, Extensions, Guides, Operations Manual, Development, Internals, and Sitemap. The main content area is titled "DuckDB Installation" and includes a version selector (Stable release v1.12.0, Nightly build), environment options (Command line, Python, R, Java, Node.js, Rust, Go, C/C++, ODBC), platform options (Windows, macOS, Linux), download methods (Package manager, Direct download), and architecture options (x86_64, arm64). The installation instructions section shows the command `winget install DuckDB.cli`. A note states that each DuckDB client is installed without relying on any other DuckDB clients. The usage example section shows the command `duckdb`.

<https://duckdb.org/>

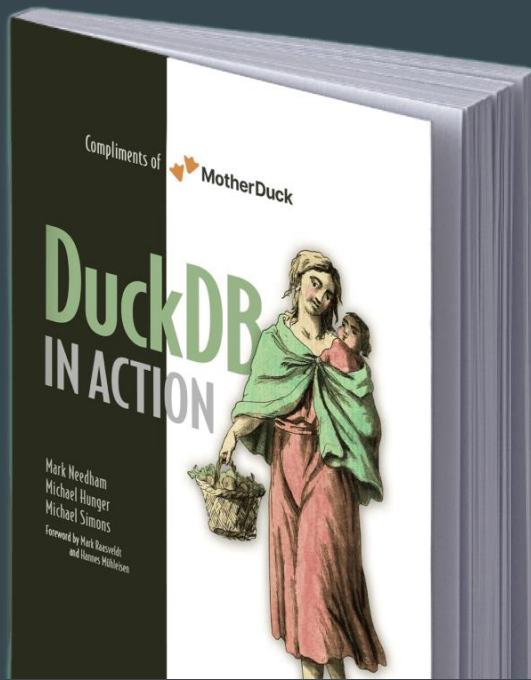


DuckDB



Data Saturday
Vitória 2024

DuckDB In Action - Mother Duck



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



DuckDB



Data Saturday
Vitória 2024

O que estão 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

db-engines.com/en/ranking

30.	↓	33.	↓	32.	Oracle Essbase	Relational	0.00	-0.02	+2.49
57.	↑	60.	↑	98.	DuckDB	Relational	5.98	+0.37	+2.34

Demetrius Mata • Seguindo
Engenheiro de Dados SR | Arquiteto de Dados Azure | AWS | Databricks | Py...
Sua dose sobre o mundo dos dados
#DuckDB #ProcessamentoDeDados #AnáliseDeDados #BigData #Tecnologia
Mundo dos Dados

Luciano Vasconcelos Filho • 1º
Engenheiro de dados e professor da Jornada
Acesse meu site
Você trabalha com SQL?
Se sim, tenho certeza que ta escutando

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

Rodrigo Santana • Seguindo
Data Engineer | Liderança em Dados e Inteligência Artificial
Aqui na empresa, implementamos o #DuckDB em alguns de nossos clientes e os resultados superaram nossas expectativas.
Na minha opinião, essas são algumas das maiores vantagens de usar o DuckDB:
• Rapidez e eficiência na execução de co

Téo Calvo • Seguindo
Eu transformo vidas por meio do ensino
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 | Aiur...
10 m •

Anderson Amaral • 1º
Building Client-Tested Wor...
4 d •

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

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

DuckDB
29,392 seguidores
10 m •

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



DuckDB



Data Saturday
Vitória 2024

O que estão 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



DuckDB



Data Saturday
Vitória 2024

Vagas?



DuckDB



Data Saturday
Vitória 2024

Caso de uso

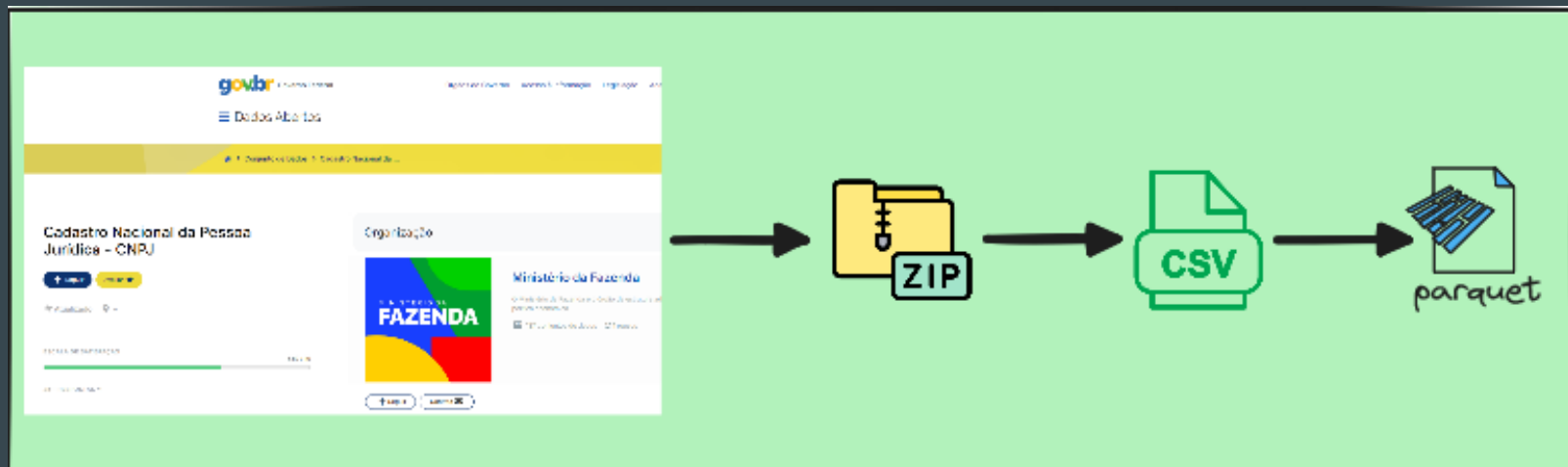
- Análise de Cadastro Nacional de Pessoas Jurídicas, dados abertos do Gov.BR;
- Apuração de concurso de rádio amador;
- Integração de sistema entre APIs/Json e Mysql;
- Documentação de pipelines Data Factory (poc);
- Migrando o curso de BD para DuckDB;



Data Saturday
Vitória 2024

Hands on

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

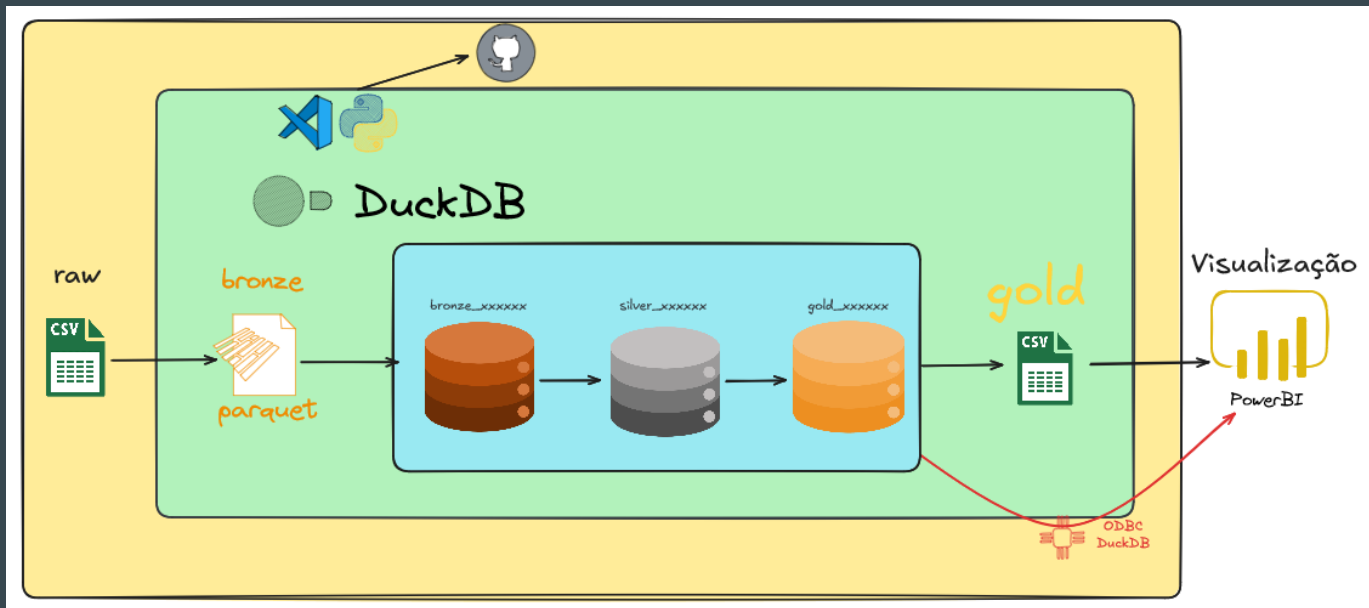


DuckDB



Data Saturday
Vitória 2024

Hands on



DuckDB



Data Saturday
Vitória 2024

Meus Contatos

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



DuckDB



Data Saturday
Vitória 2024

Arquivos

https://github.com/profdanilosantos/duckdb_data_saturday_2024_vitoria



DuckDB



Data Saturday
Vitória 2024