

Análise de Transparência Orçamentária e Contratual com DuckDB e Python: Um Estudo de Caso da Gestão COVID-19 em Natal (2020)

Cayo César da Costa Lopes
Gabriel Gomes dos Santos
Gustavo Douglas Silva Cruz
Joseph Douglas Moura Batista
Joyce Oliveira dos Santos

Departamento de Computação e Tecnologia
Universidade Federal do Rio Grande do Norte

Julho de 2025

DuckDB

Sumário

- O que é o DuckDB
- Em que se destaca
- Conceitos iniciais
- Comandos básicos
- Instalação do ambiente
- Demonstração prática

O que é o DuckDB

O DuckDB é um Sistema de Gerenciamento de Banco de Dados (SGBD) moderno, projetado para análise de dados rápida e eficiente.

Diferente de bancos tradicionais como PostgreSQL ou MySQL, que seguem um modelo cliente-servidor, o DuckDB opera como um banco de dados embutido, sendo executado diretamente dentro do seu aplicativo, sem a necessidade de um servidor dedicado.

Isso o torna particularmente eficaz para o processamento rápido e eficiente de grandes volumes de dados locais, especialmente quando utilizado em conjunto com linguagens como Python.

Em que o DuckDB se destaca

■ Analítica (OLAP):

- Otimizado para consultas analíticas em grandes volumes.
- Armazenamento em colunas para maior desempenho.

■ Simplicidade e Portabilidade:

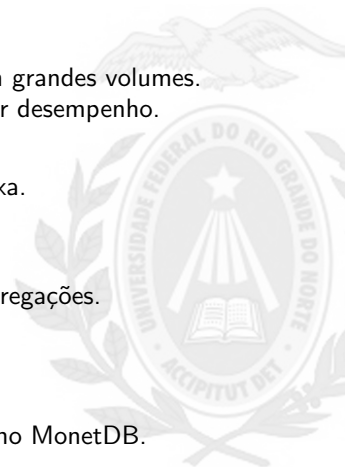
- Biblioteca leve, sem instalação complexa.
- Lê diretamente CSV, Parquet e JSON.

■ SQL Avançado e Extensível:

- Suporte a window functions, joins e agregações.
- Permite extensões e leitura de APIs.

■ Open Source com Base Acadêmica:

- Licença MIT, livre para uso.
- Fundado em pesquisas de sistemas como MonetDB.



Arquitetura Analítica (OLAP)

■ Alta performance em OLAP:

- Otimizado para consultas analíticas em grandes volumes de dados.
- Armazenamento em colunas, ideal para agregações e filtrações.
- Execução vetorizada: processa dados em blocos com alta eficiência.

■ Simplicidade e Portabilidade:

- Funciona como biblioteca em Python, R, C++, etc.
- Dispensa instalação complexa: basta importar.
- Lê arquivos CSV, Parquet e JSON diretamente.
- Compatível com notebooks Jupyter e aplicações web (via DuckDB-Wasm).

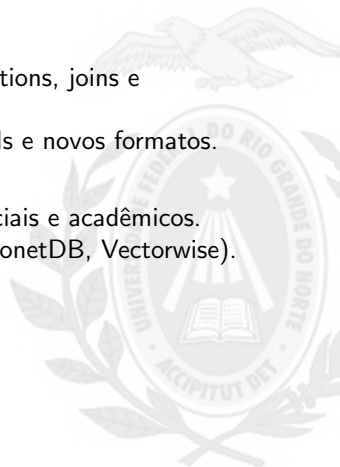
SQL e Comunidade

■ SQL Completo e Extensível:

- Suporte a SQL avançado: window functions, joins e agregações.
- Permite extensões, como leitura de APIs e novos formatos.

■ Open Source e Comunidade Ativa:

- Licença MIT: uso livre para fins comerciais e acadêmicos.
- Baseado em pesquisas consolidadas (MonetDB, Vectorwise).



Comandos Básicos no DuckDB

Criação de tabela e inserção:

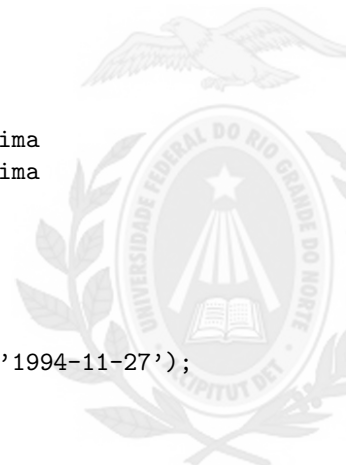
```
CREATE TABLE weather (  
    city    VARCHAR,  
    temp_lo INTEGER, -- temperatura mínima  
    temp_hi INTEGER, -- temperatura máxima  
    prcp    FLOAT,  
    date    DATE  
);
```

```
INSERT INTO weather  
VALUES ('San Francisco', 46, 50, 0.25, '1994-11-27');
```

Consultas:

```
SELECT * FROM weather;
```

```
SELECT DISTINCT city FROM weather;
```



Comandos Básicos no DuckDB (continuação)

Atualização de dados:

```
UPDATE weather
SET temp_hi = temp_hi - 2,
    temp_lo = temp_lo - 2
WHERE date > '1994-11-28';
```

Remoção de registros:

```
DELETE FROM weather
WHERE city = 'Hayward';
```



Instalação do DuckDB

Linux (via terminal):

```
curl https://install.duckdb.org | sh
```

Windows (baixar CLI):

```
<https://github.com/duckdb/duckdb/releases/download/v1.3.1/  
duckdb_cli-windows-amd64.zip>
```

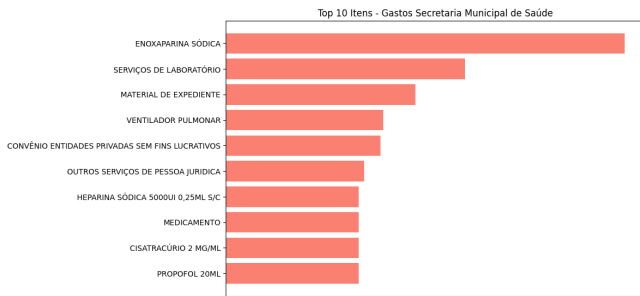
Python (via pip):

```
pip install duckdb --upgrade
```



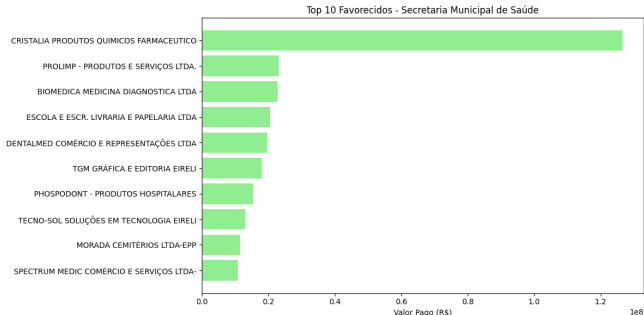
Gastos por Item na Secretaria Municipal de Saúde

- Analisamos os principais itens que consumiram recursos financeiros.
- Os 10 itens com maiores valores pagos indicam onde houve maior investimento ou despesa.
- Essa visão ajuda a identificar áreas prioritárias e possíveis oportunidades de otimização.



Gastos por Favorecido na Secretaria Municipal de Saúde

- Identificamos os 10 maiores favorecidos, ou seja, os principais recebedores dos pagamentos.
- Essa análise é importante para transparência e para compreender a distribuição dos recursos.
- Permite monitorar fornecedores e parceiros estratégicos.



Referências

- <https://duckdb.org/>
- https://coronavirus.natal.rn.gov.br/transparencia/dados_abertos.html

