

# Duckdb: Embedded Analytics DB of Future

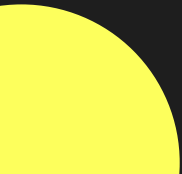
Course

---

Developed by: Maxim Migutin



This Course will be  
**continuously updated**  
based on students'  
**feedback & feature**  
**requests**

A yellow circular graphic is located in the bottom-left corner of the slide.

# What will you learn?

**Features &  
Capabilities**



**Practice/  
Hands-on cases**

# What will you learn?

**DuckDB  
Theory**

**Competition**

**Market Niche**

**Typical Use  
Cases**

**Setting up &  
Configuration**

**CLI usage**

**SQL  
Innovations**

**Python  
integration**

**Streamlit:  
Analytical Apps**

**Data  
Engineering:  
dbt**

**MotherDuck**



01

# Course Intro



# duckdb: What & Why?

## 1. What?

**A modern embedded analytics db that runs locally and crunches GIGABYTES of data easily**

## 2. Why?

**Fast & Fun Analytical research without the need for bulky Infra (Hadoop, Spark, etc.)**

# How fast is DuckDB?



## Real World Stress Test:

- 10M+ rows
- 5 test SQL cases
- DuckDB vs Pandas vs Polars
- 1 MacBook Air with 16gb RAM

Technology	Time taken (sec)	Runtime increase vs. DuckDB
Duckdb 🔥	4.9	-
polars	6.49	+32%
pandas	21.84	+345%

# How fast is DuckDB?



benchmark.py ...

[100%]

benchmark: 3 tests										
Name (time in s)	Min	Max	Mean	StdDev	Median	IQR	Outliers	OPS	Rounds	Iterations
test_benchmark_duckdb	3.8859 (1.0)	4.2227 (1.0)	4.0326 (1.0)	0.1725 (1.05)	3.9892 (1.0)	0.2526 (1.09)	1;0	0.2480 (1.0)	3	1
test_benchmark_polars	5.6241 (1.45)	5.9344 (1.41)	5.7470 (1.43)	0.1649 (1.0)	5.6826 (1.42)	0.2327 (1.0)	1;0	0.1740 (0.70)	3	1
test_benchmark_pandas	18.1832 (4.68)	19.8782 (4.71)	19.0131 (4.71)	0.8480 (5.14)	18.9779 (4.76)	1.2712 (5.46)	1;0	0.0526 (0.21)	3	1

## Legend:

Outliers: 1 Standard Deviation from Mean; 1.5 IQR (InterQuartile Range) from 1st Quartile and 3rd Quartile.

OPS: Operations Per Second, computed as 1 / Mean

3 passed in 150.49s (0:02:30)



2

# DuckDB: Introduction

What's so special about this  ?

## 2.1 What is duckdb?

**duckdb** = modern embedded OLAP db:

- can run locally & in the Cloud
- rapidly growing ecosystem

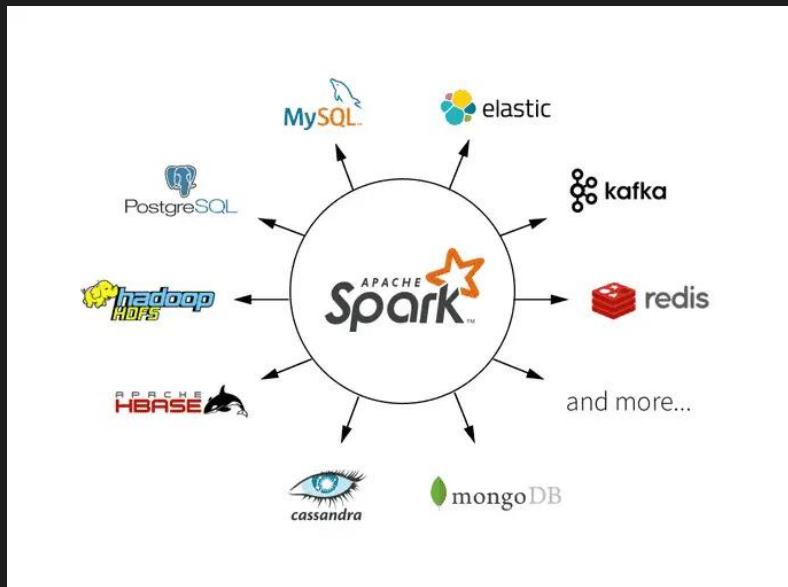


## 2.2 What's “under the hood”? 🤔

1. Column storage & Columnar-vectorized query execution engine ⚙️
2. No external dependencies & No “Client-Server” architecture: just an embedded process 📦
3. Single-file databases (just like SQLite but OLAP-oriented) 💾
4. Feature-rich SQL dialect (tabular functions, window functions) 🤖
5. Seamless Python & R integration 🤝
6. Support of advanced data structures (arrays, structs, maps) 🔬

## 2.3 Why should you use duckdb? 💪

# 1. No need for “bulky” Big Data Infra unless you “really need it” 🍔



vs

● DuckDB

## 2. Easy set up compared to traditional RDBMS 🧠



AWS Redshift

VS



DuckDB



Clickhouse



*cassandra*

### 3. Faster than SQLite for Analytics-related tasks\*



vs



*\*while also on-device deployable*

#### 4. Faster & needs less resources than Pandas 🚀

Pandas

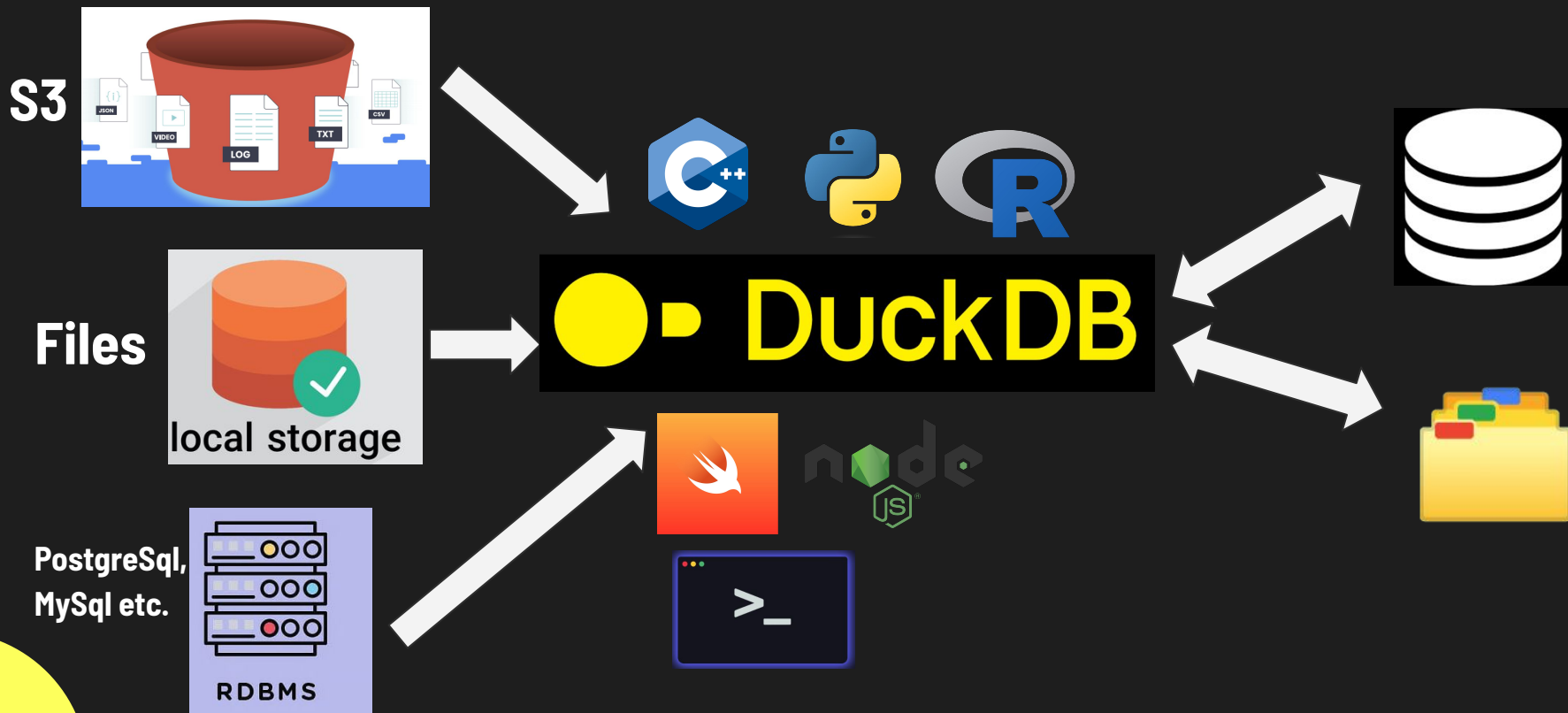


vs












## 2.4.1 Where does duckdb stand?








## 2.4.2 Duckdb's competition & niche

	Transactional	Analytical
Standalone	 	 
Embedded	 	

## 2.5 When should you use duckdb?

- “Background” analytical processing for existing Apps (Cloud/Mobile/CLI)
- Data Science/Analytics workflows on 1 machine
- Cross-platform data analysis: Edge + Cloud + Local
- Replacing AWS Athena/Redshift pipelines with cheaper DuckDB-powered cloud functions

## 2.6 Who is duckdb for?

1. **Data Scientists/Analysts & ML Engineers** 
2. **Data Engineers** 
3. **DevOps Specialists** 
4. **IoT Developers** 
5. **Software Engineers** 



3

# Environment set up & DEMO

Installing duckdb & functionality kick-off



4

# duckdb CLI

Exploring duckdb's SQL innovations



5

# duckdb & Python

Integrating duckdb into DS/DA/DE workflows



6

# duckdb & Streamlit

Creating Analytical Apps just became easier!



7

# duckdb & DBT



Streamlining Data Engineering



8

# Motherduck



Cloud-native SaaS for duckdb