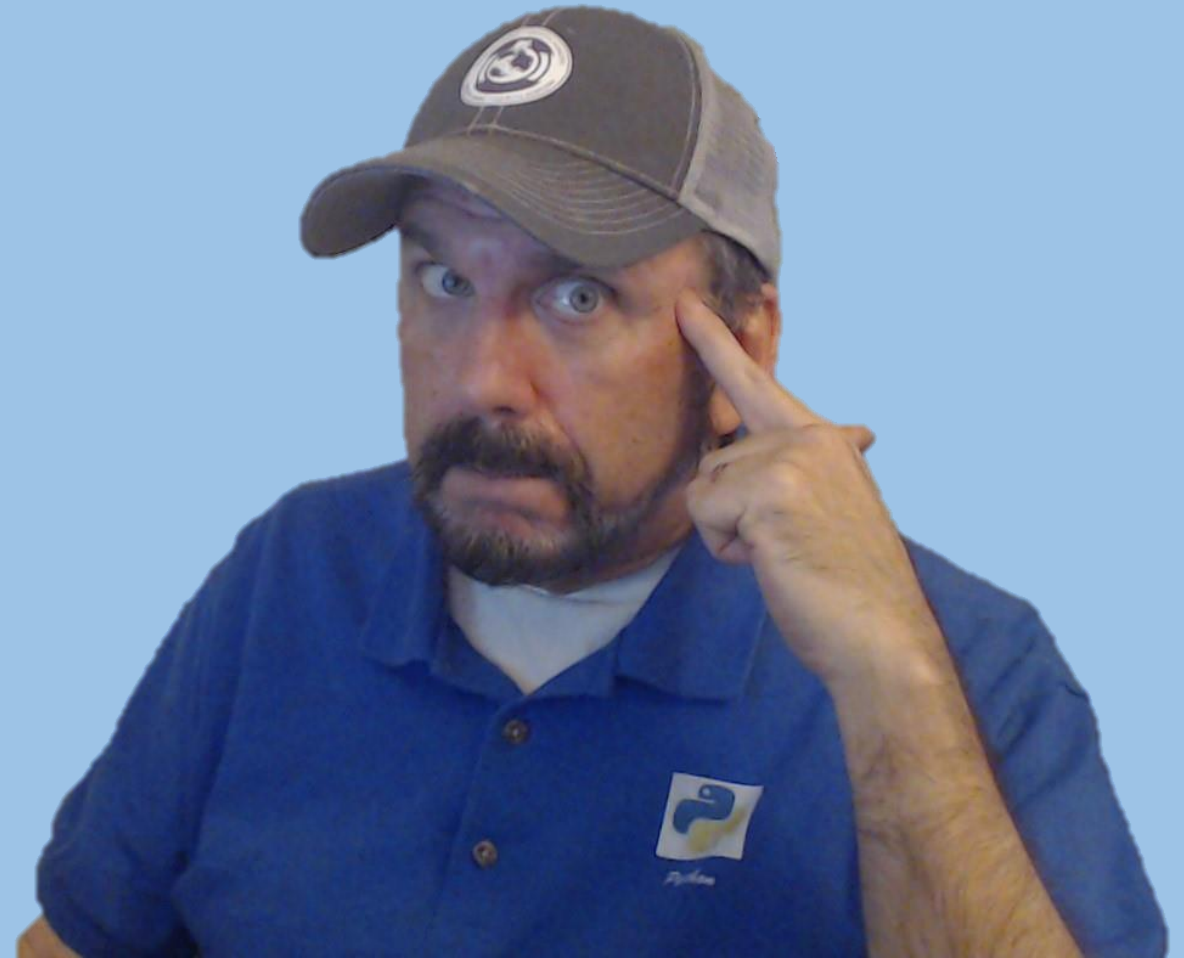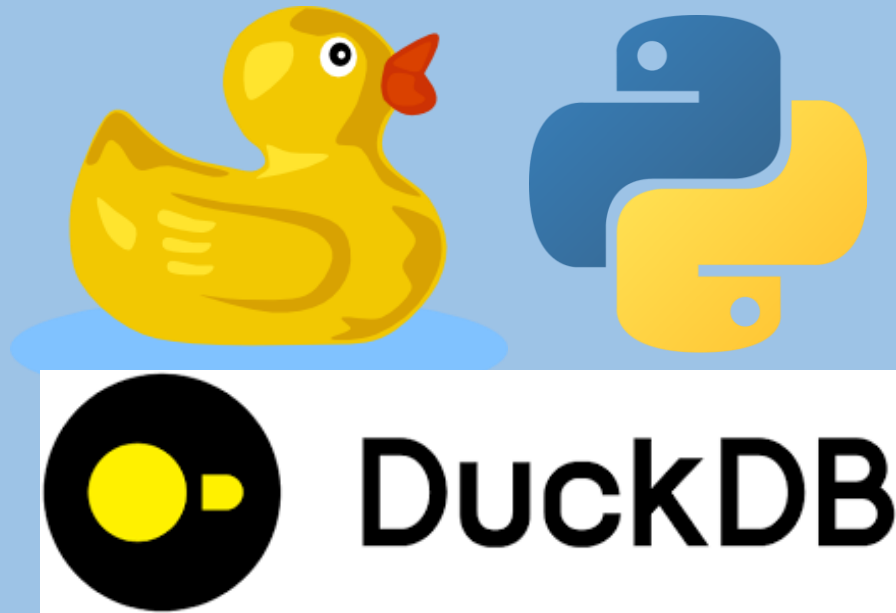# Quick Intro to DuckDB using Python

Where Are We Going?

➤ What is DuckDB?

➤ Why use DuckDB?
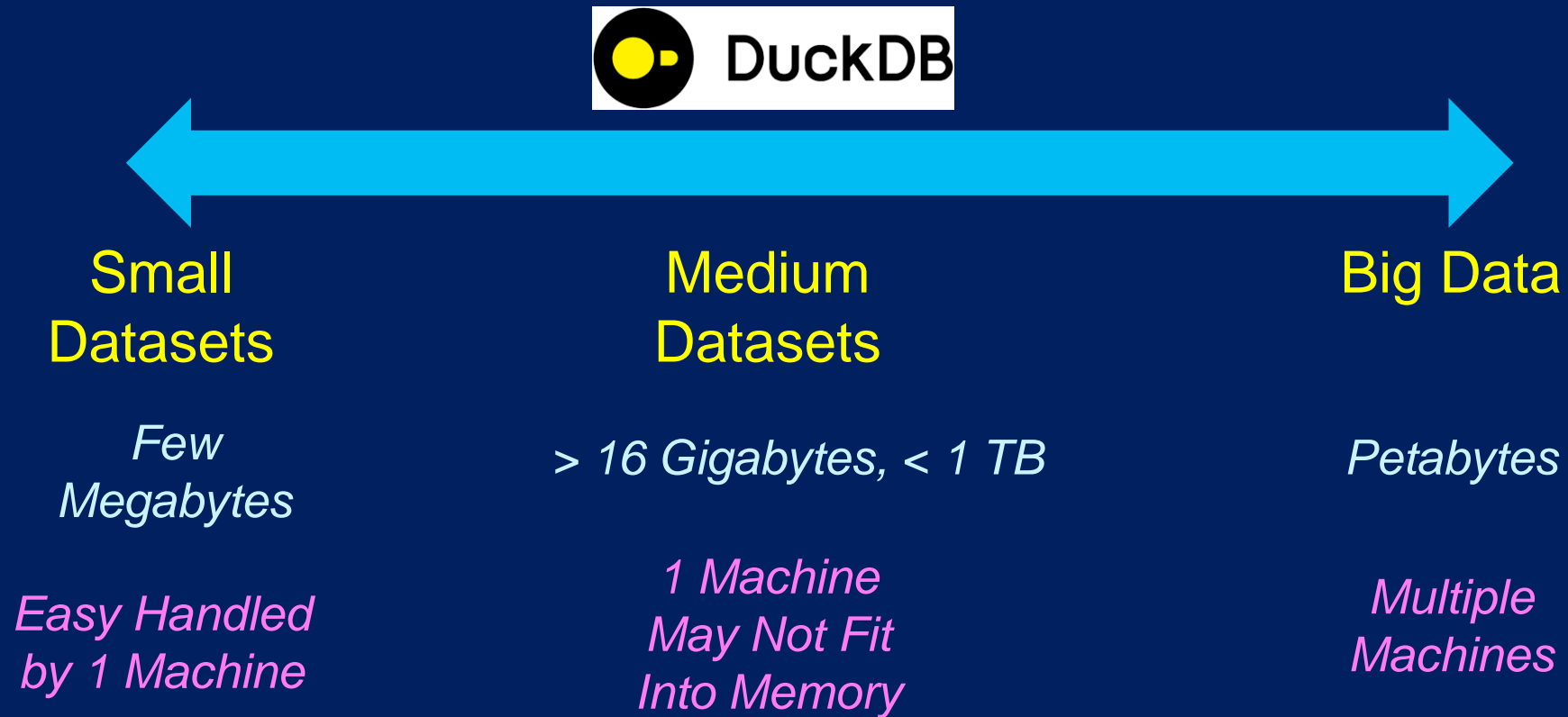
➤ Challenges

➤ Demo

What is DuckDB?

- A Database that Runs In-Process like SQLite

- Uses Column Store Format.

- Saves Tables to a Flat File or You Can Work in Memory.

- Provides a Database Service & Full SQL Interface.
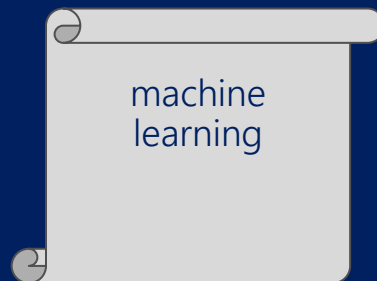
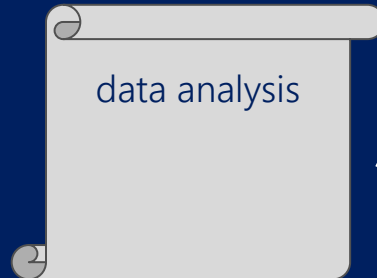- Basically, DuckDB is like an open-source Tabular Model.

https://duckdb.org/

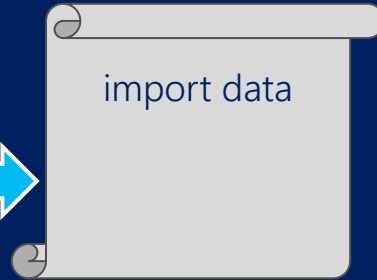# Supporting the Range of Data Volume

DuckDB

Small Datasets

Few Megabytes

Easy Handled by 1 Machine

Medium Datasets

> 16 Gigabytes, < 1 TB

1 Machine May Not Fit Into Memory

Big Data

Petabytes

Multiple Machines

DuckDB Maximizes the Single Machine

Laptop

External Data

import data

data analysis

machine learning

Local File
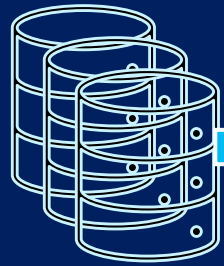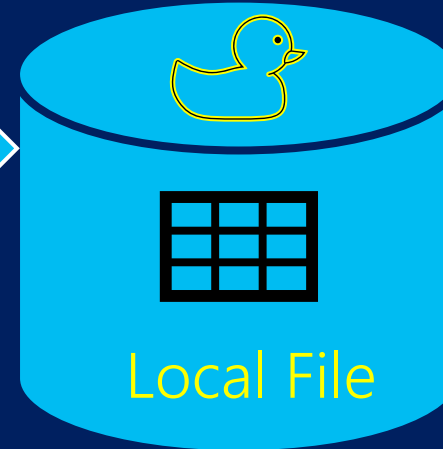
DuckDB

c:\data\myduckdb.db

Why Use DuckDB?

➢ It's Very Fast Especially with Large Datasets

➢ Handles Data that Does Not Fit Into the Machine's Memory

➢ Super Easy to Install & No Dependencies

➢ Free & Open Source – Liberal License

➢ Full SQL Support

➢ Support Many Languages or Stand Alone

Learning Challenges

➢ The Different APIs Are Not Clearly Distinguished

  ❖ Python API & SQL API (and many more) – Covered Independently

➢ Concepts and Objects Are Not Always Intuitive

➢ Documentation Examples Are Limited

Demo

Wrapping Up

➢What is DuckDB?

➢Why use DuckDB?

➢Challenges

Thank You!