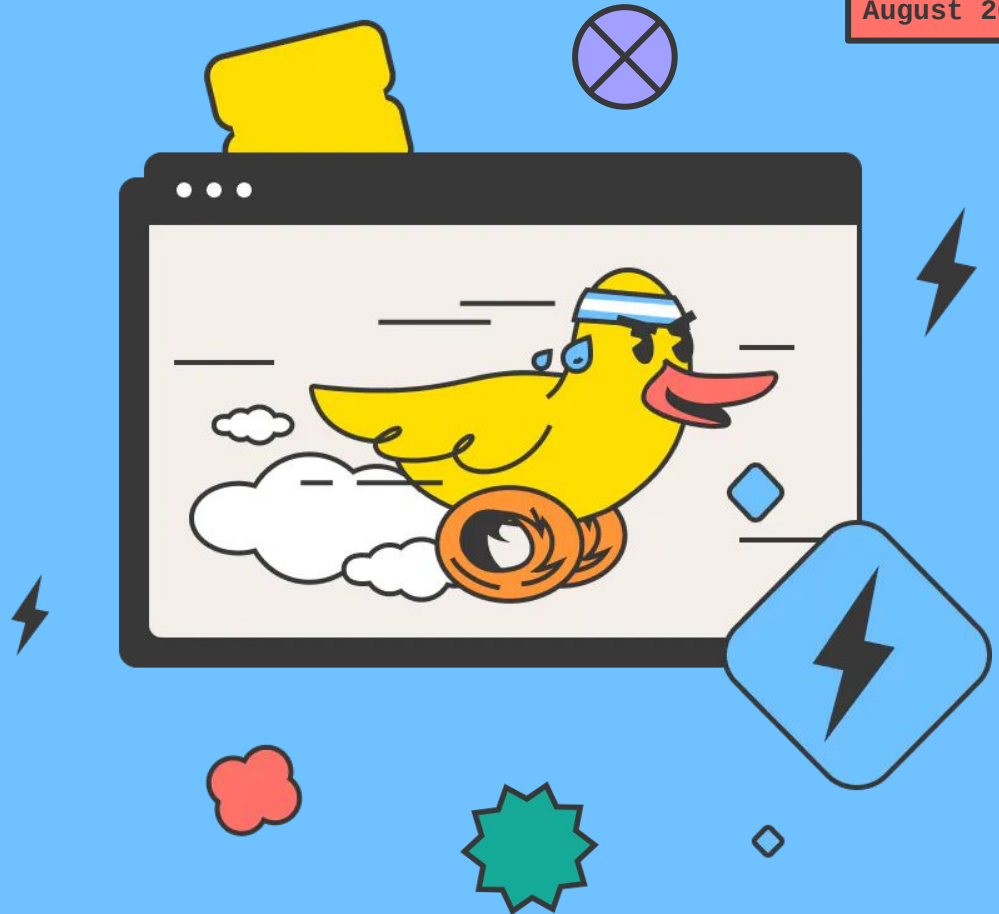


DuckCon #5  
August 2024

# TAKING FLIGHT WITH INTERACTIVE ANALYTICS



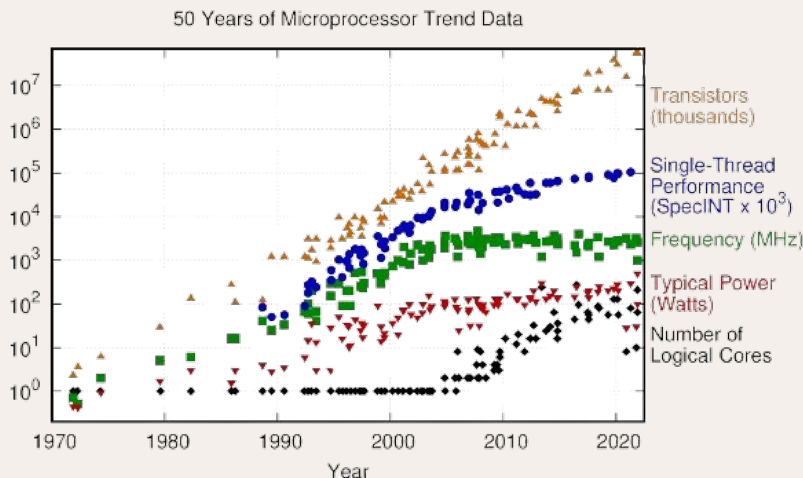
FRANCES PERRY  
 MotherDuck

Big Data is Dead,  
Long Live Small Data



# Compute Economics Have Fundamentally Shifted

More *cores*, instead of more *machines*.



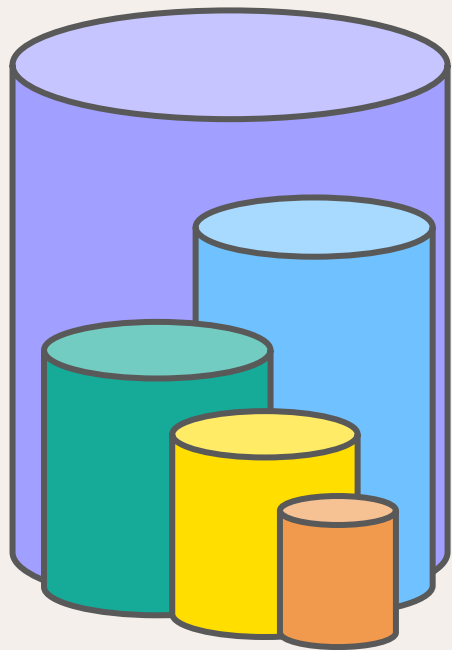
Original data up to the year 2010 collected and plotted by M. Horowitz, F. Laborte, O. Shacham, K. Okotun, L. Hamrond, and C. Batten  
New plot and data collected for 2010-2021 by K. Rupp

**2004:** MapReduce ran on **dual-processor** x86 processors running Linux, with **2-4 GB** of memory per machine.

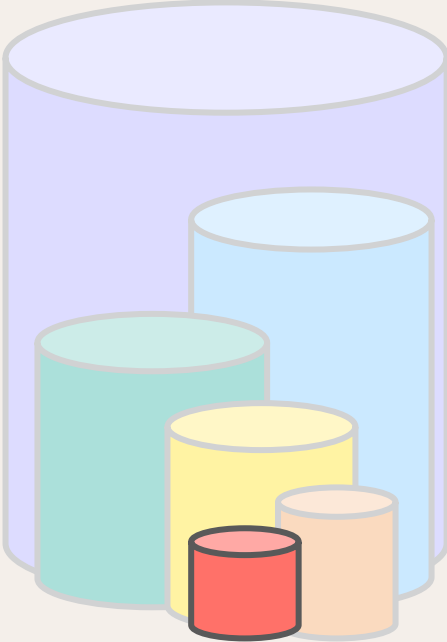
**2024:** Google Cloud's general purpose C4 machines have up to **96 cores** and **1.5TB** of memory.

**This Laptop:** The Macbook air running my slides has **8 cores** and **16GB** of ram

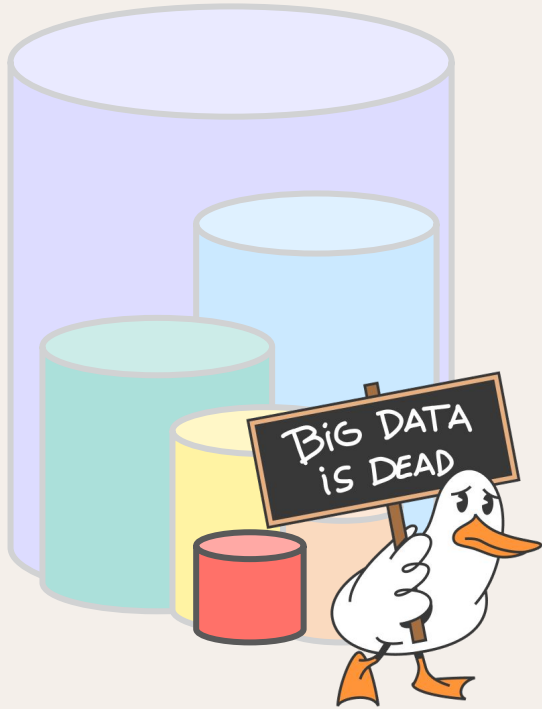
Even if we accumulate Big Data...



Even if we accumulate Big Data...



# Big Data is Dead



**Most people don't actually have big data.**

**People who have it query small tables anyway.**

**People who do query big data only read a small bit of it.**

**Big data tools are for the "Big Data 1%".**

# Big Data is Dead: The Redshift Files



**Most people don't actually have big data.**

✓ 95% of databases aren't big.

**People who have it query small tables anyway.**

✓ When people have big data, 99% of their queries use smaller tables.

**People who do query big data only read a small bit of it.**

✓ 99.5% of queries over big data tables, query small data amounts.

**Big data tools are for the "Big Data 1%".**

✓ And actually, even just the Big Data 0.00025%.

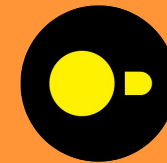


**DuckDB**





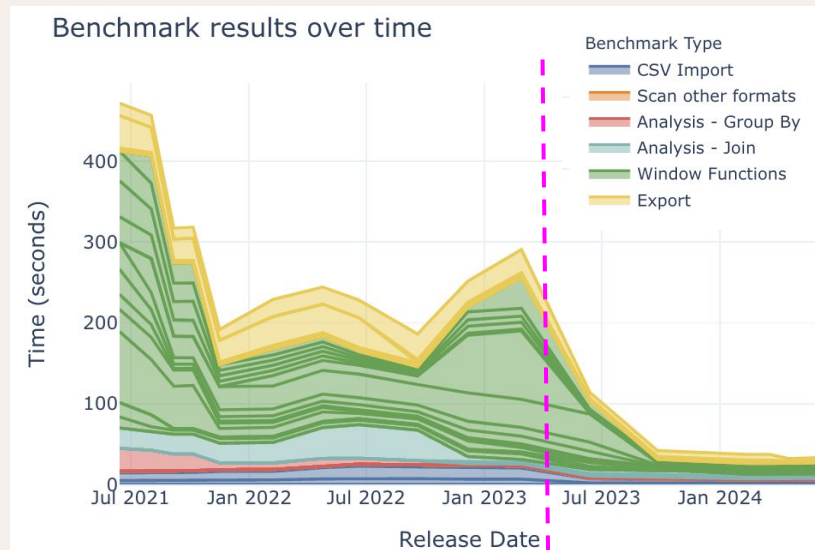
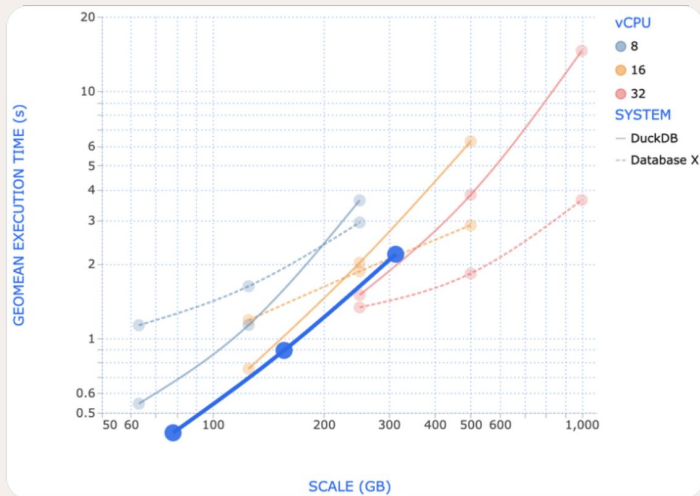
# Increasingly Performant



George Fraser @frasergeorgew · Mar 21 2023

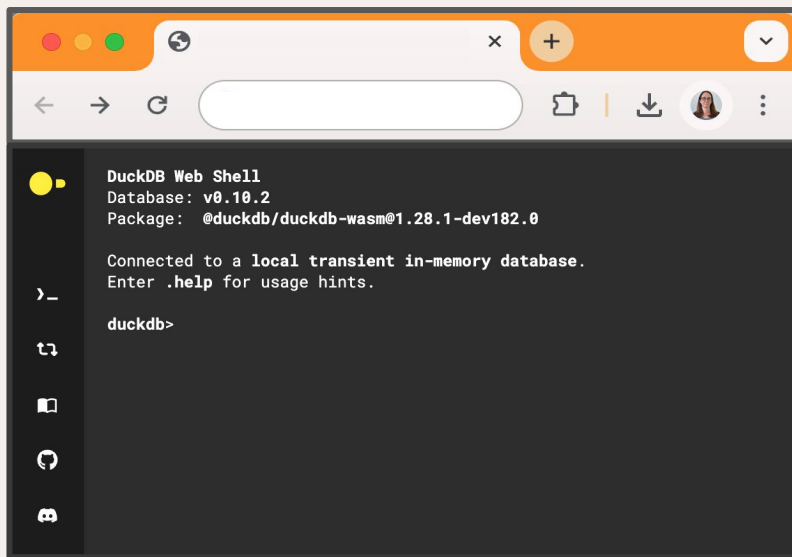
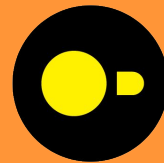
Replying to @frasergeorgew

But my most astonishing finding: the M1 Pro is an absolute beast. The laptop in front of me is faster than most of the systems I benchmarked!



March 2023

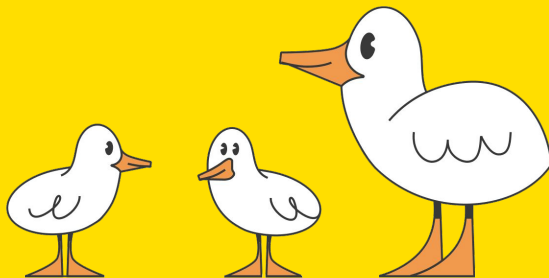
# Freedom to Run Anywhere



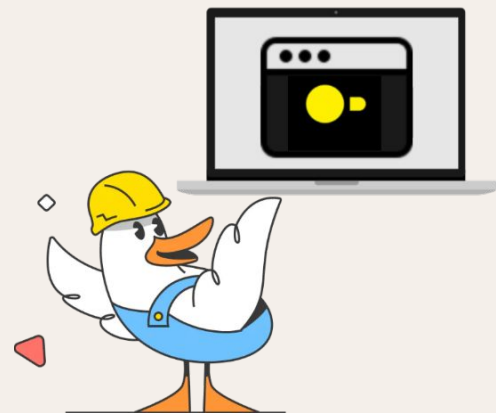
- Zero external dependencies
- Runs embedded within a host process
- Compiles into WASM, allowing it to run directly in any major web browser.



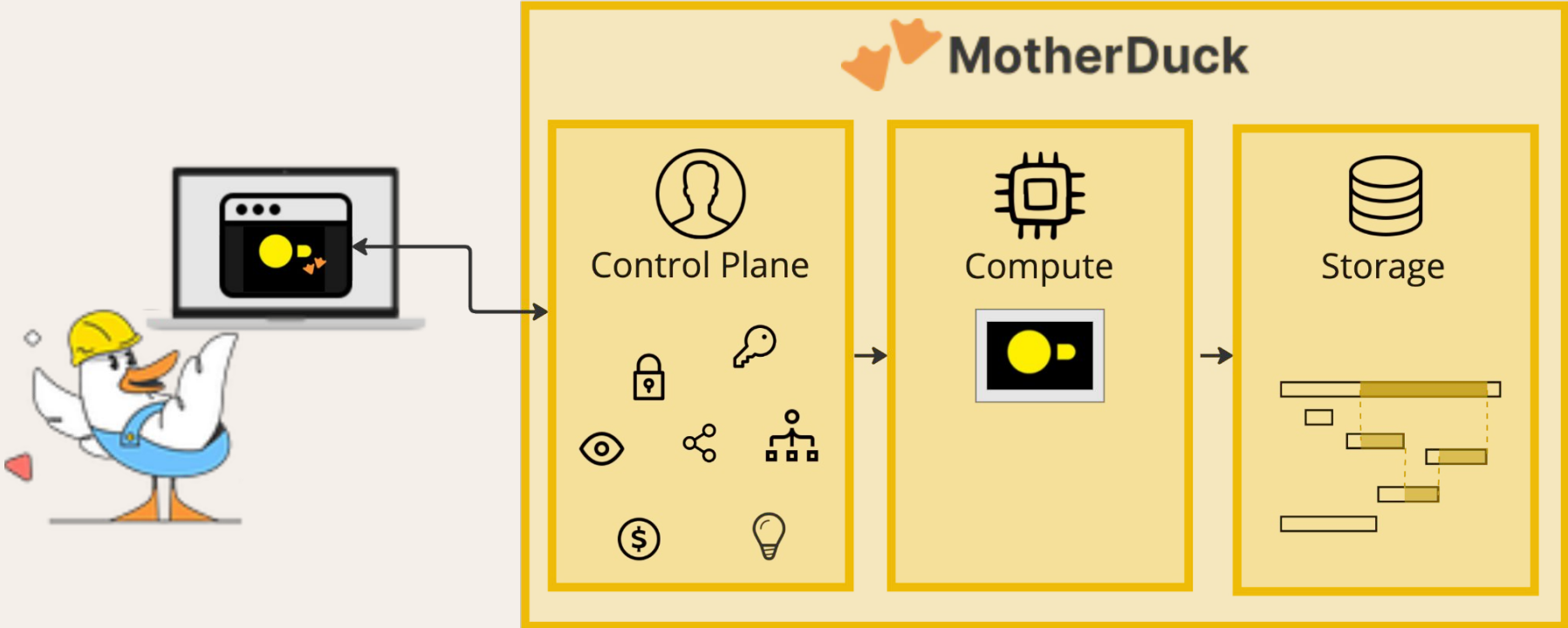
# Mother Duck



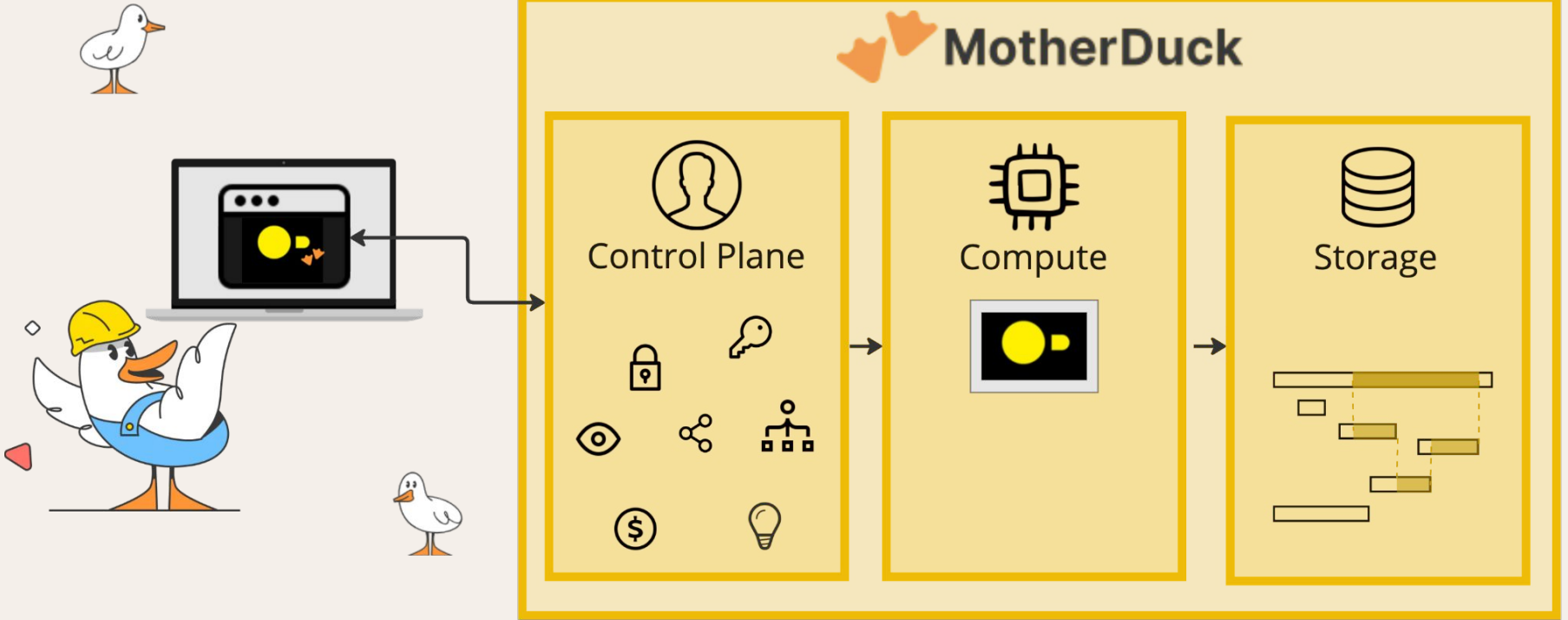
# Managed, Multiplayer DuckDB



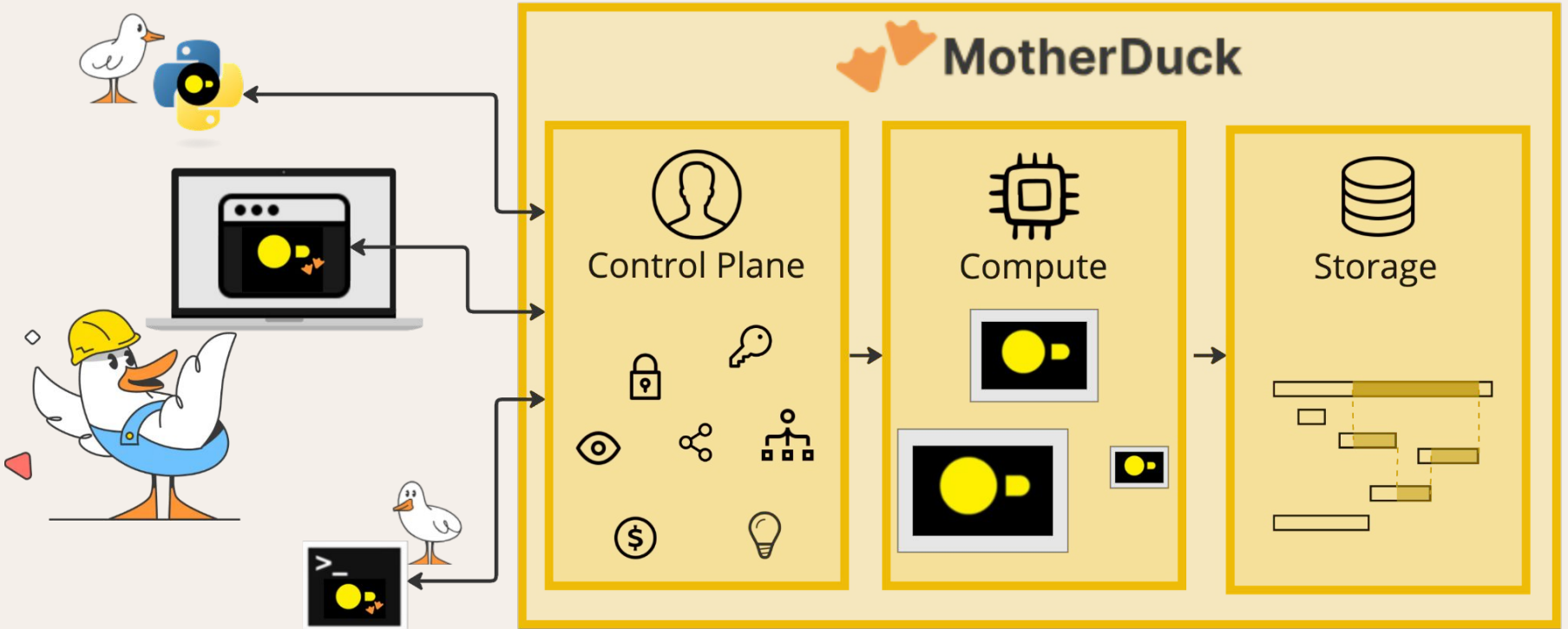
# Managed, Multiplayer DuckDB



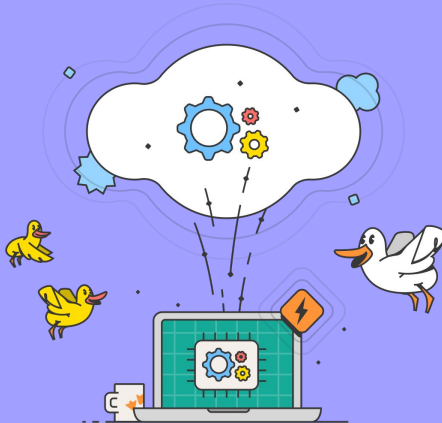
# Managed, Multiplayer DuckDB



# Managed, Multiplayer DuckDB

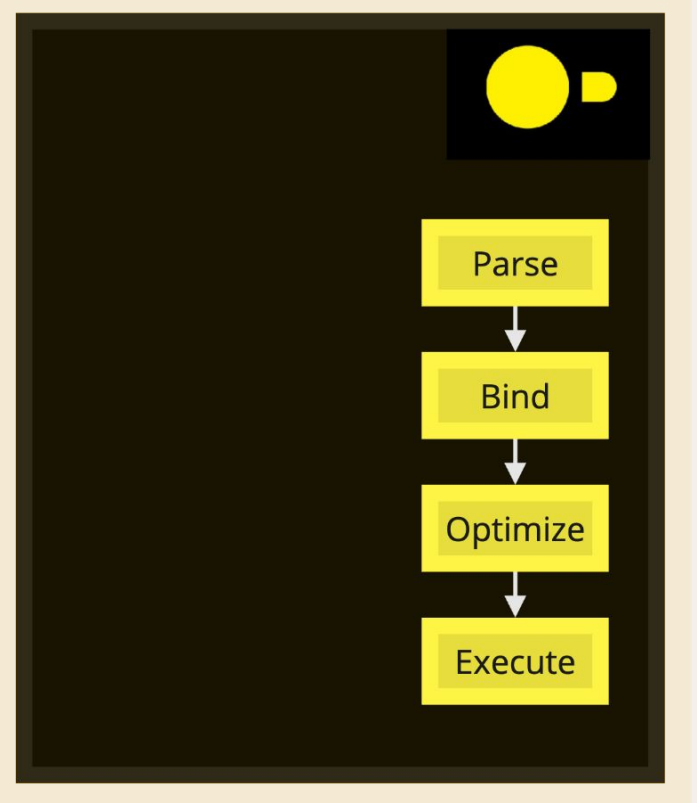
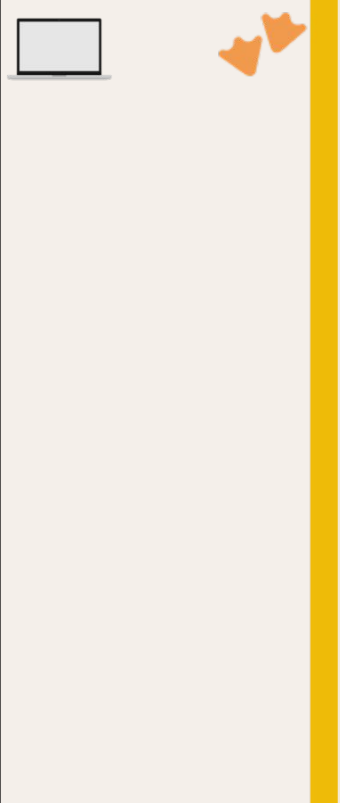
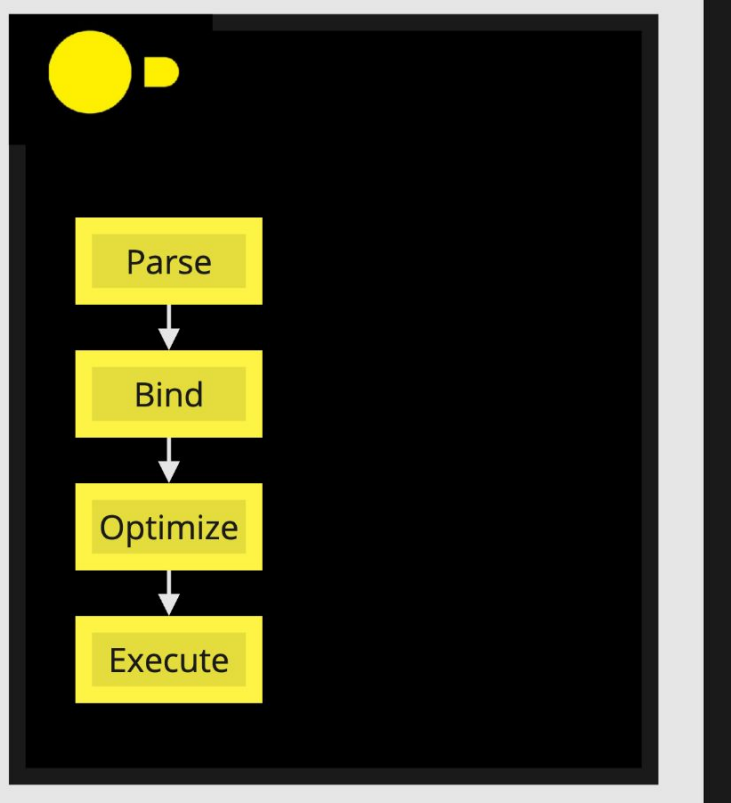


# Dual Execution

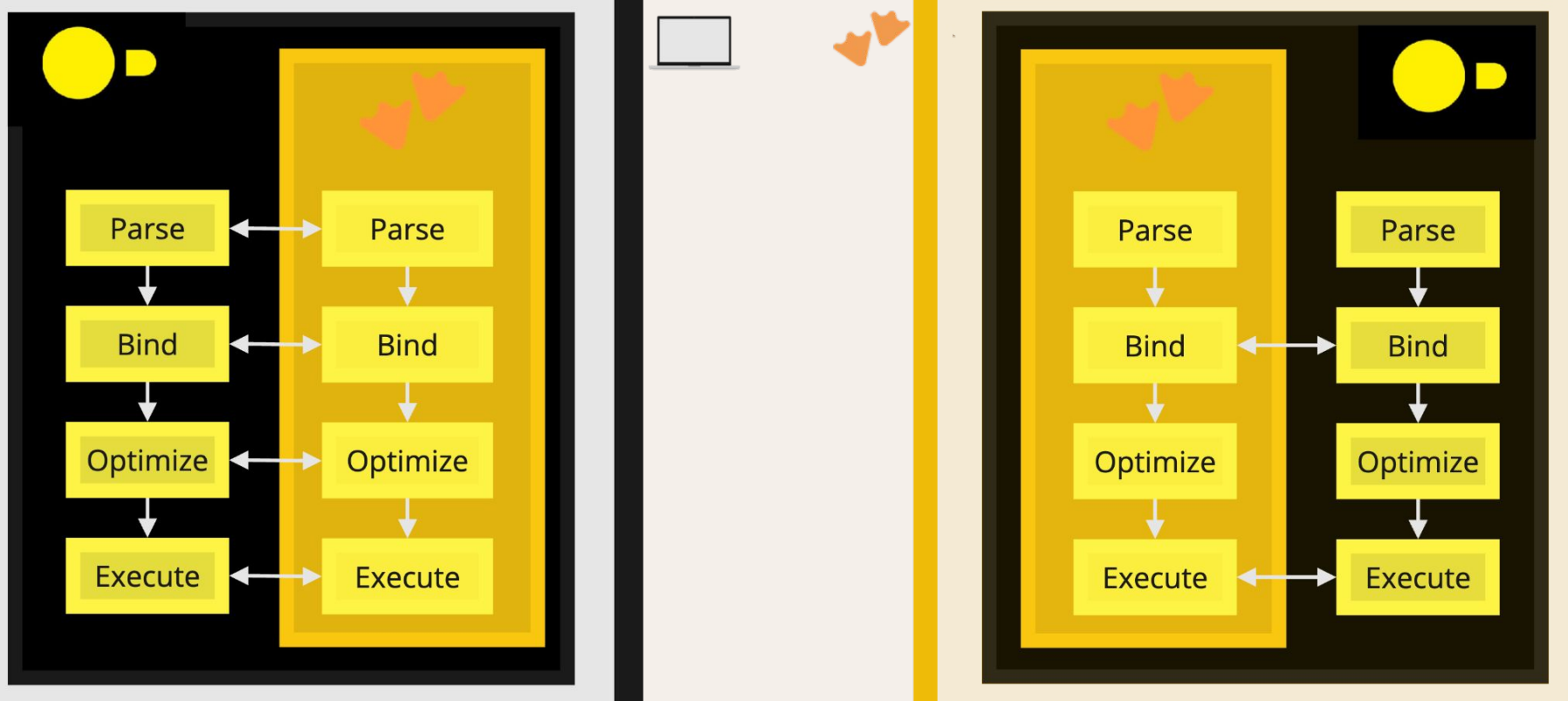




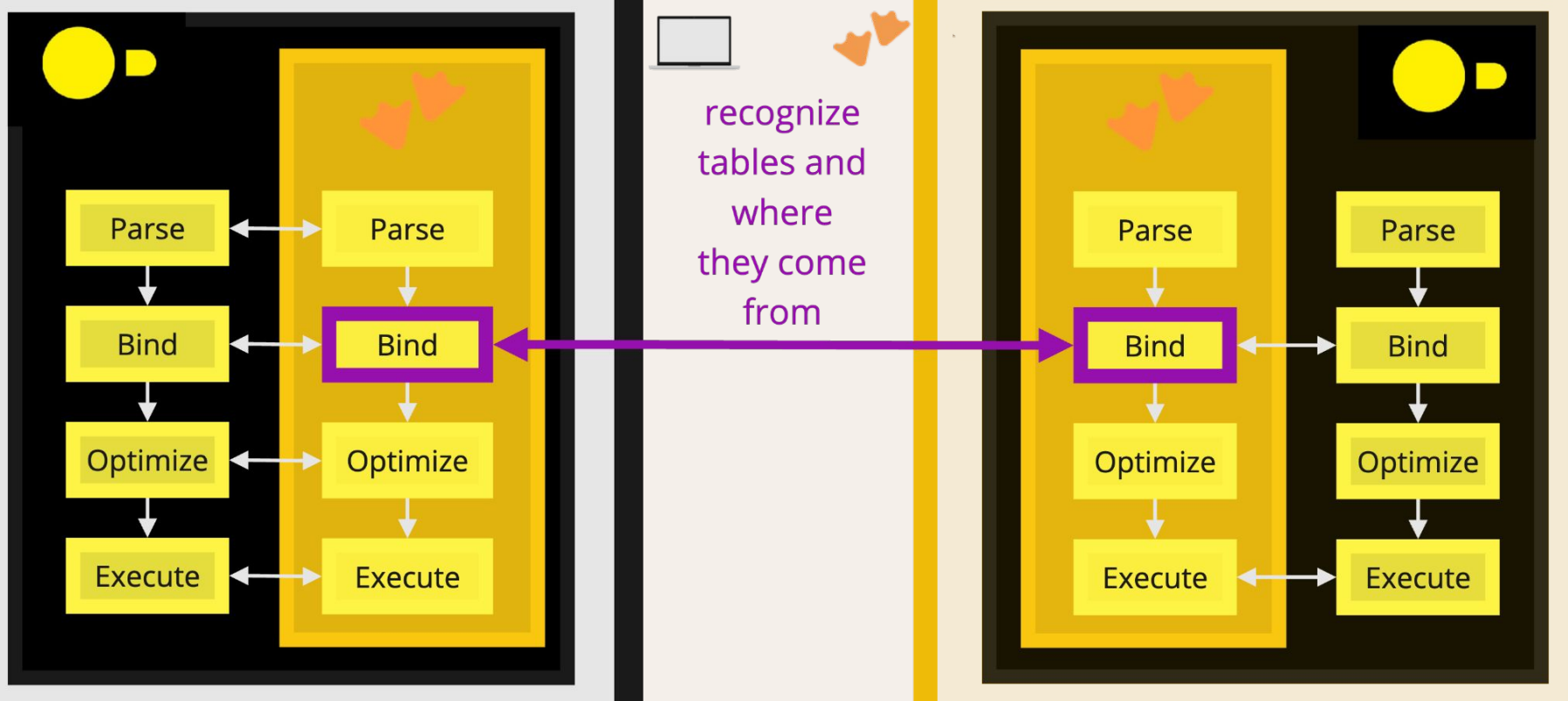
# Dual Execution Architecture



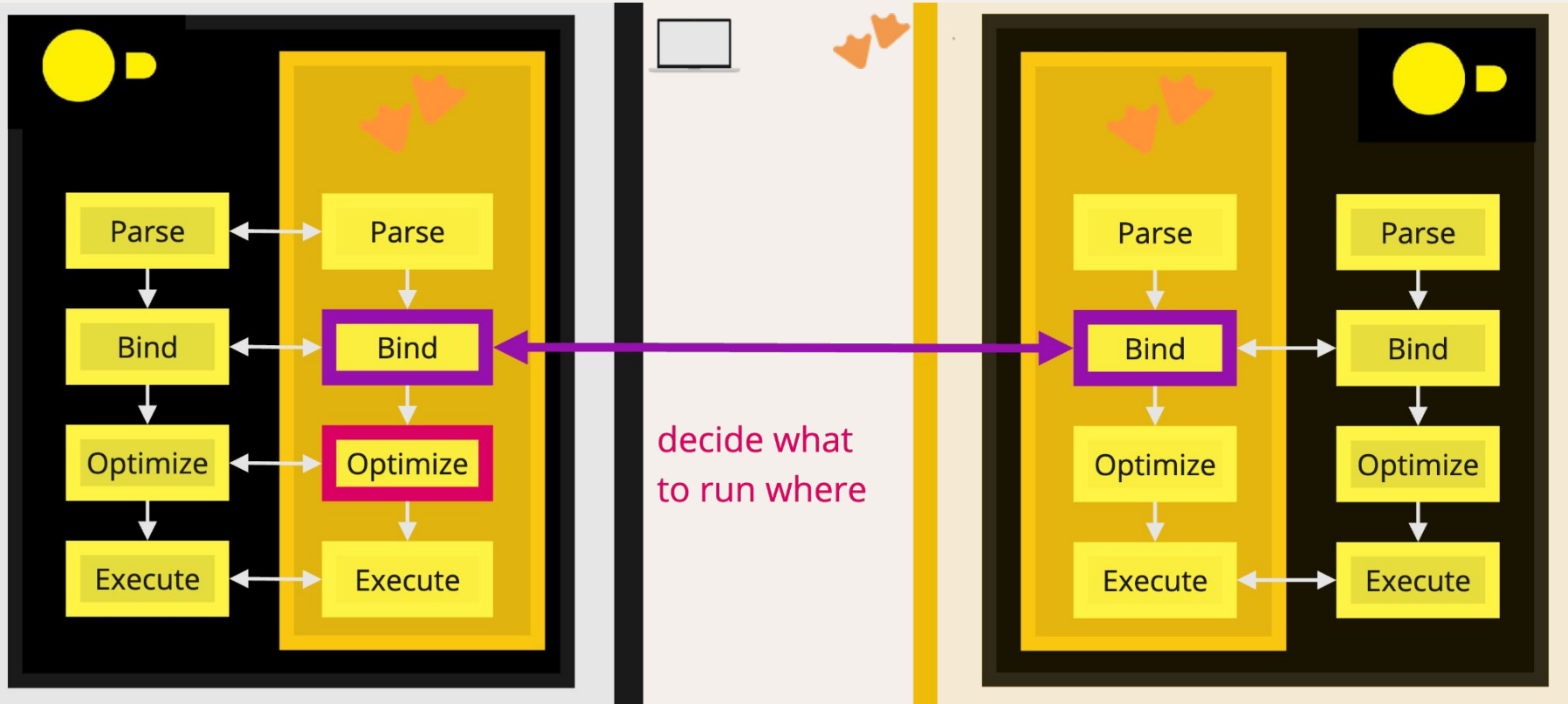
# Dual Execution Architecture



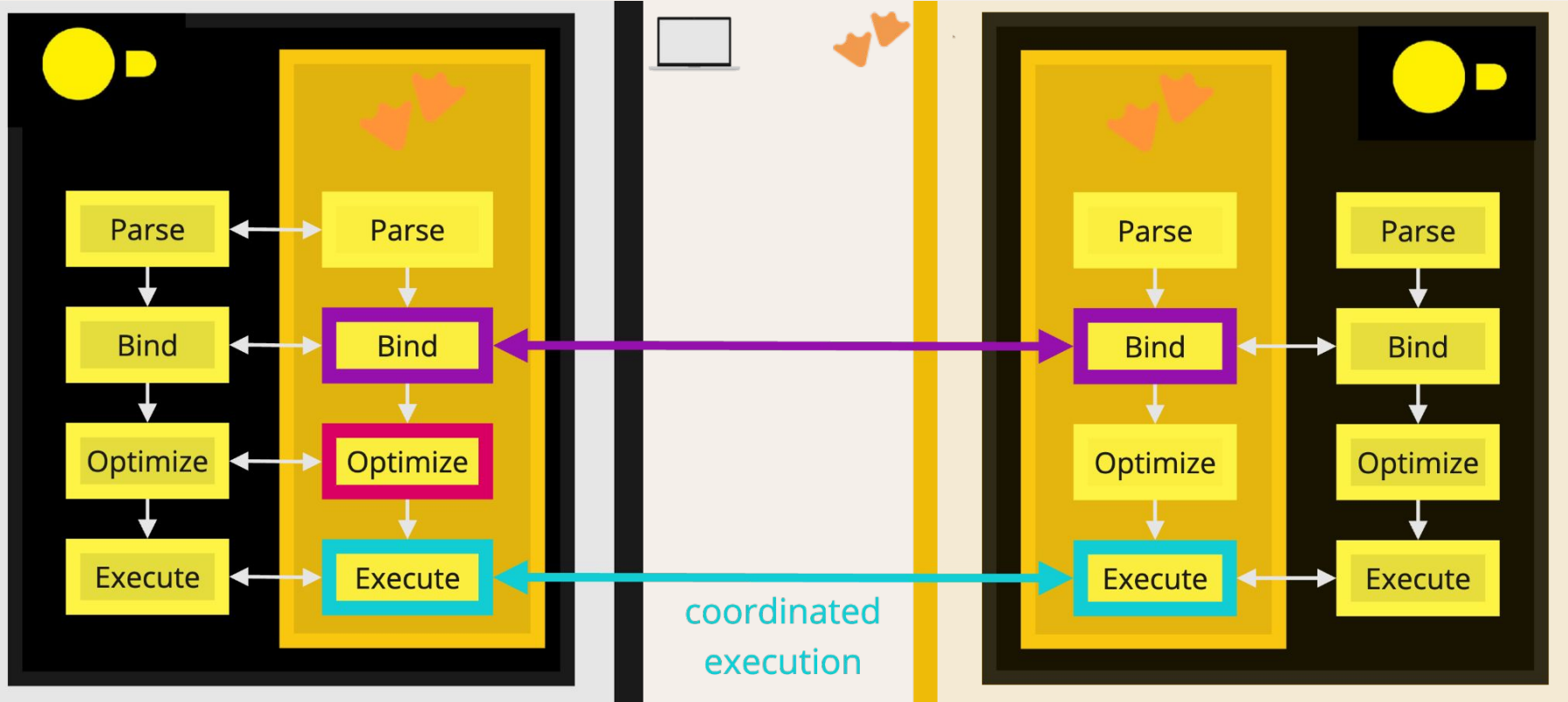
# Dual Execution Architecture



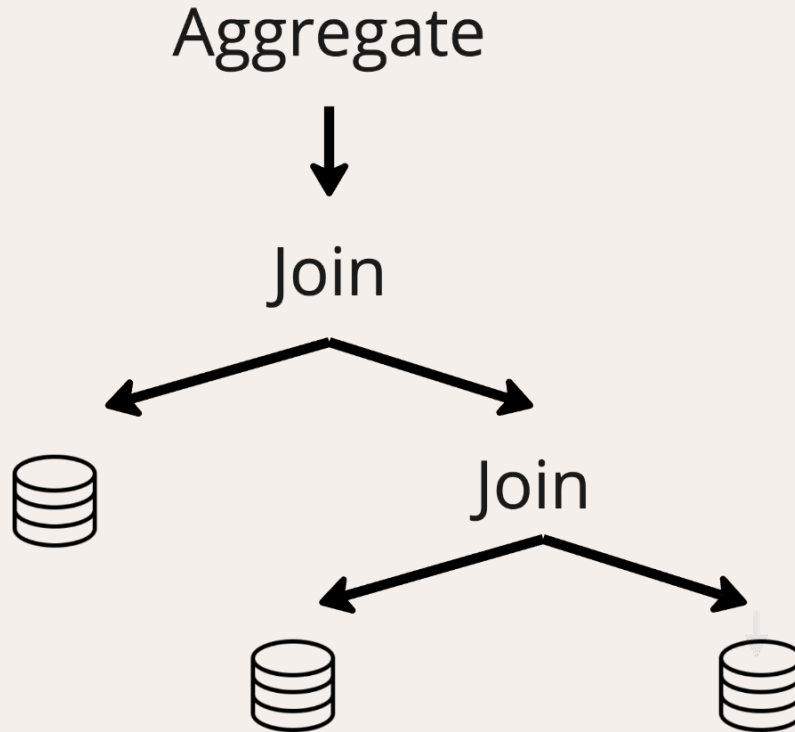
# Dual Execution Architecture



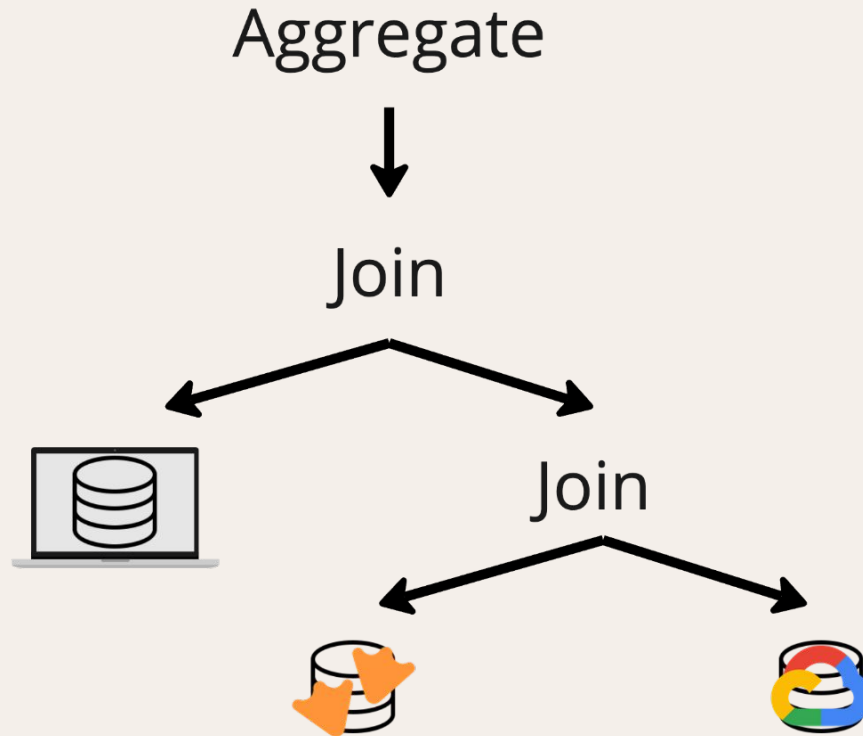
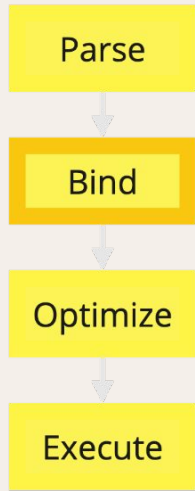
# Dual Execution Architecture



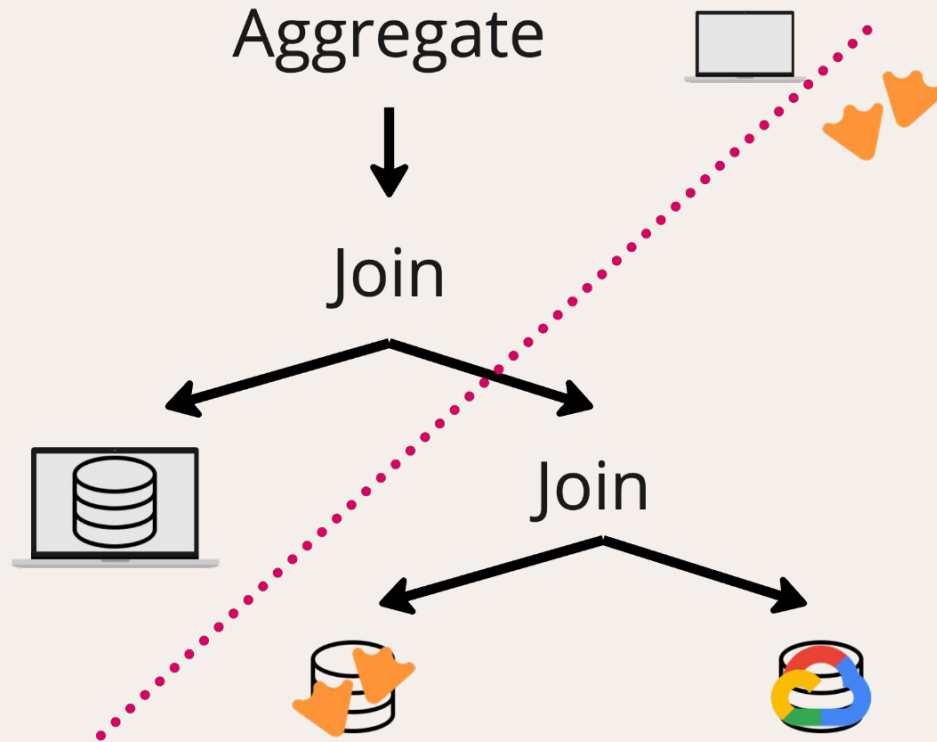
# Dual Execution Example



# Dual Execution Example

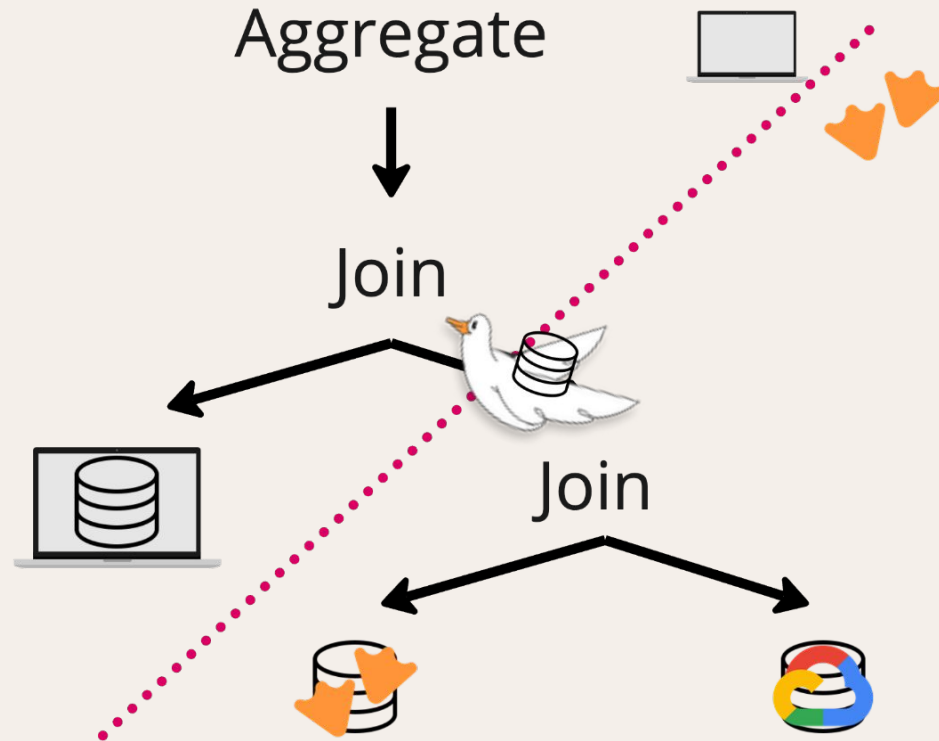


# Dual Execution Example

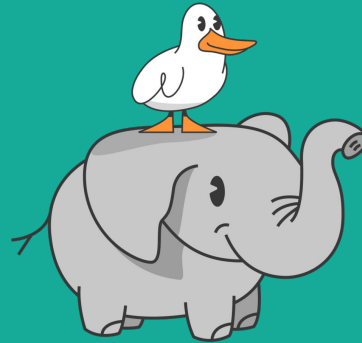
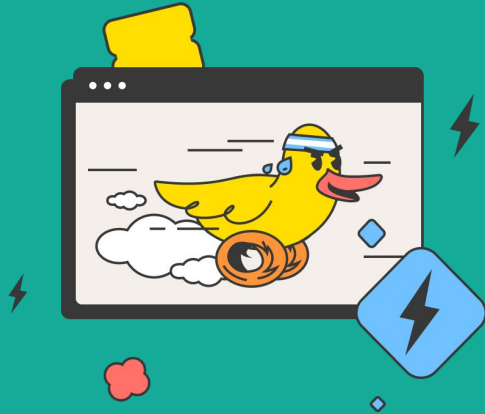




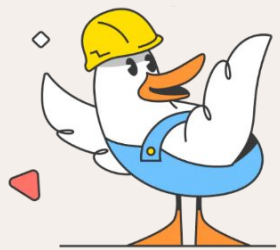
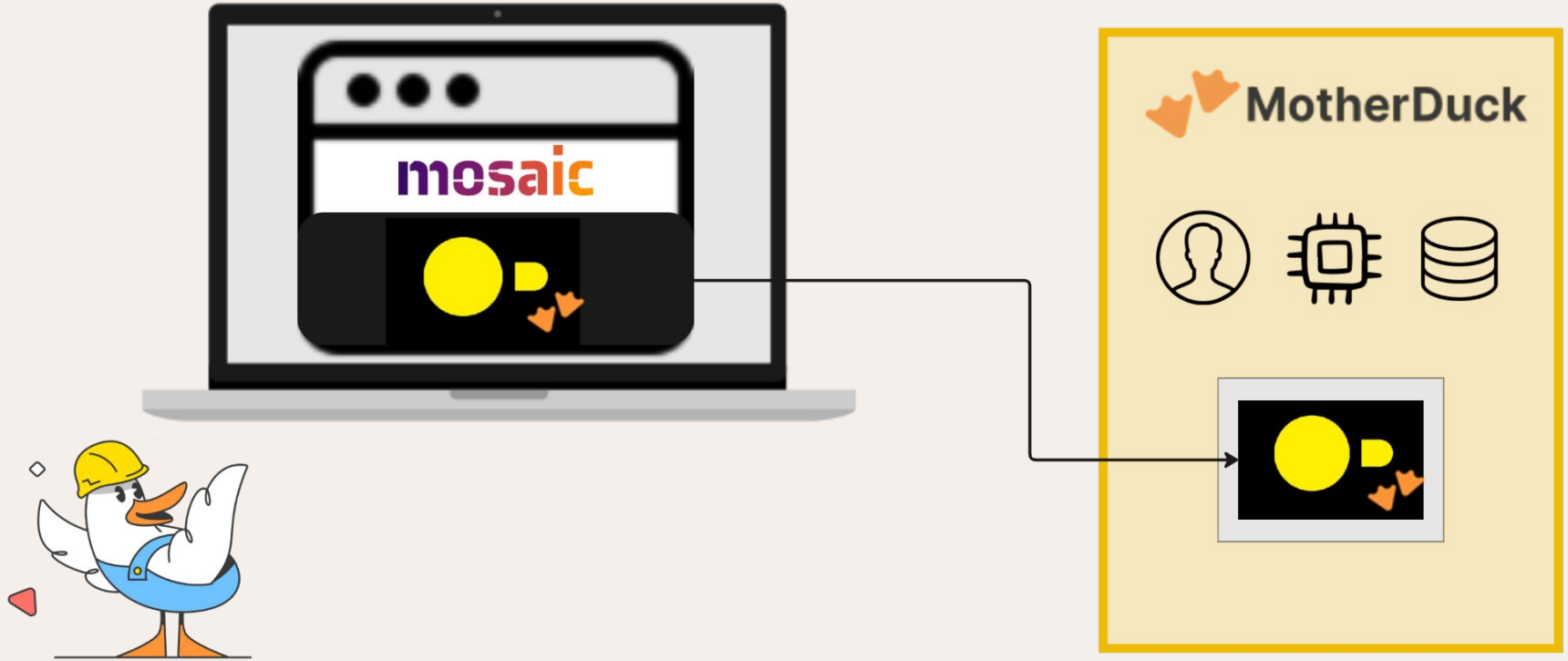
# Dual Execution Example



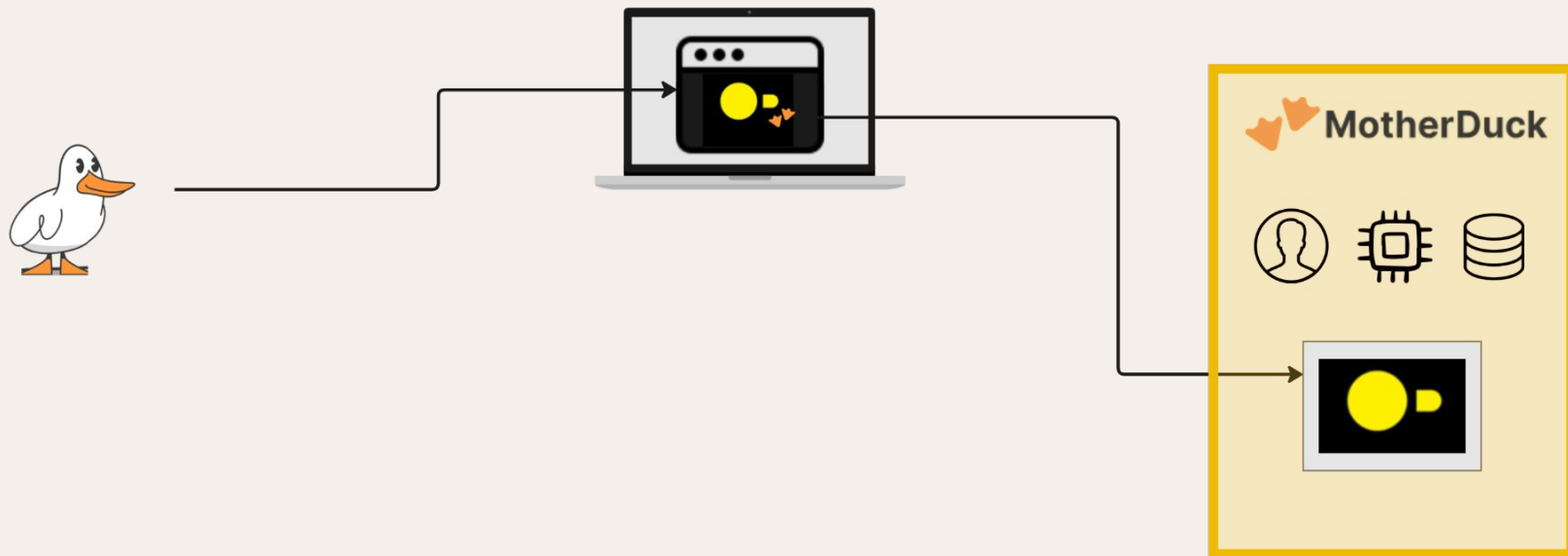
# Demo Time



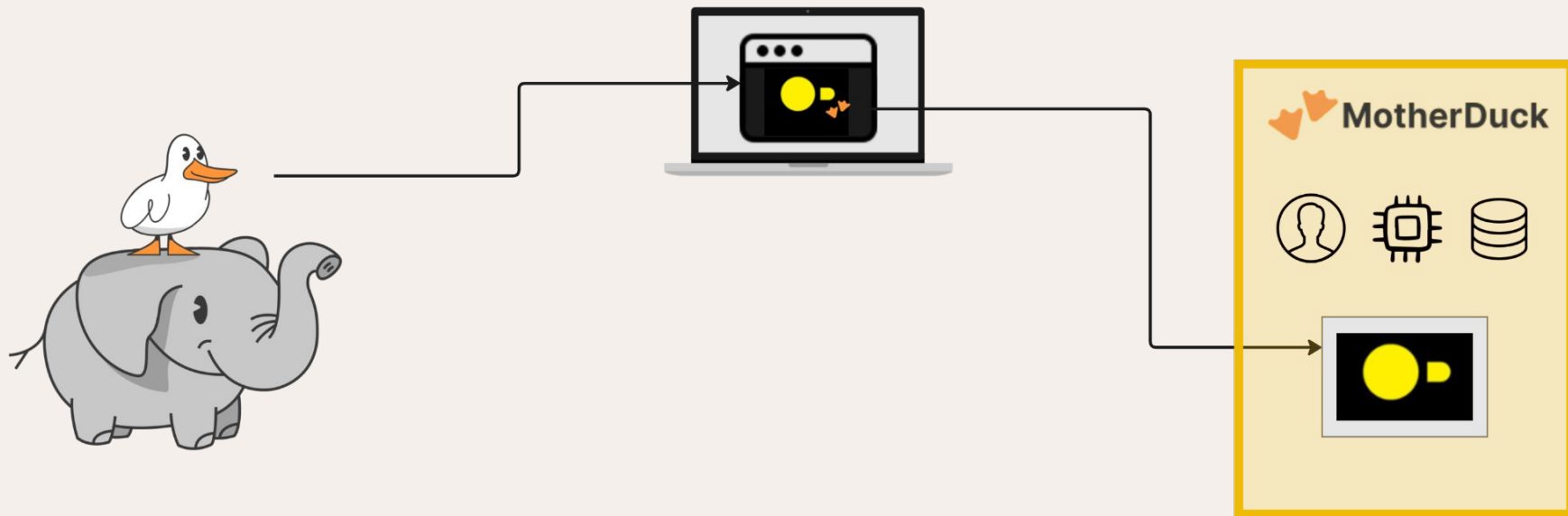
# Instantaneous, Interactive Visualizations



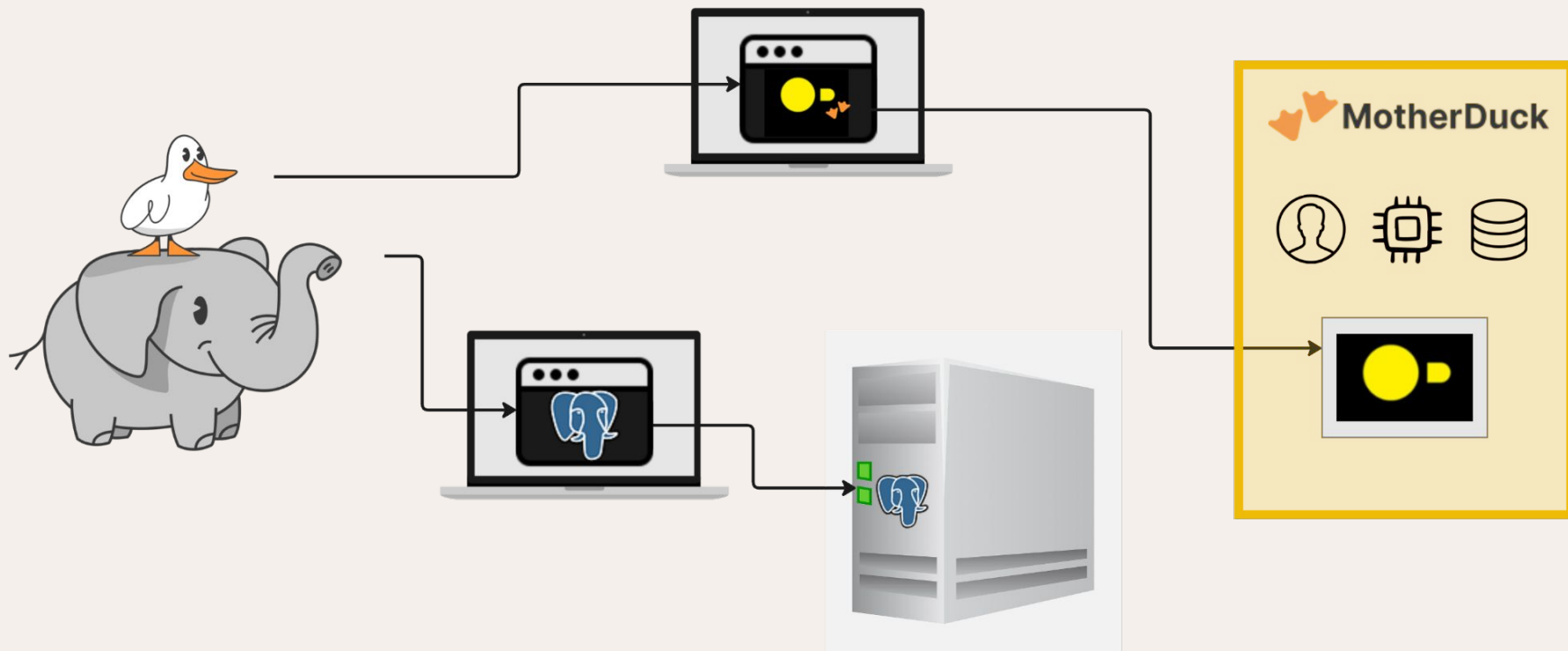
# Embedded Analytical Processing



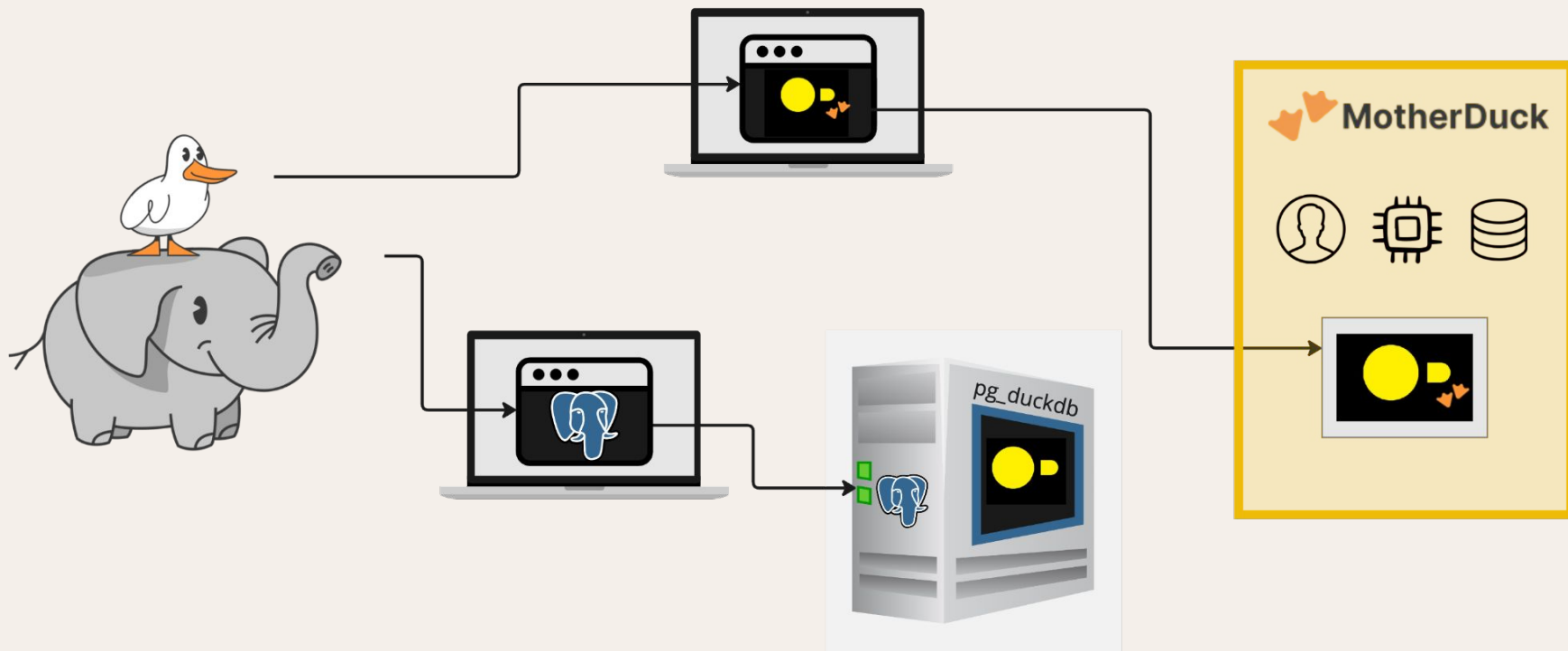
# Embedded Analytical Processing



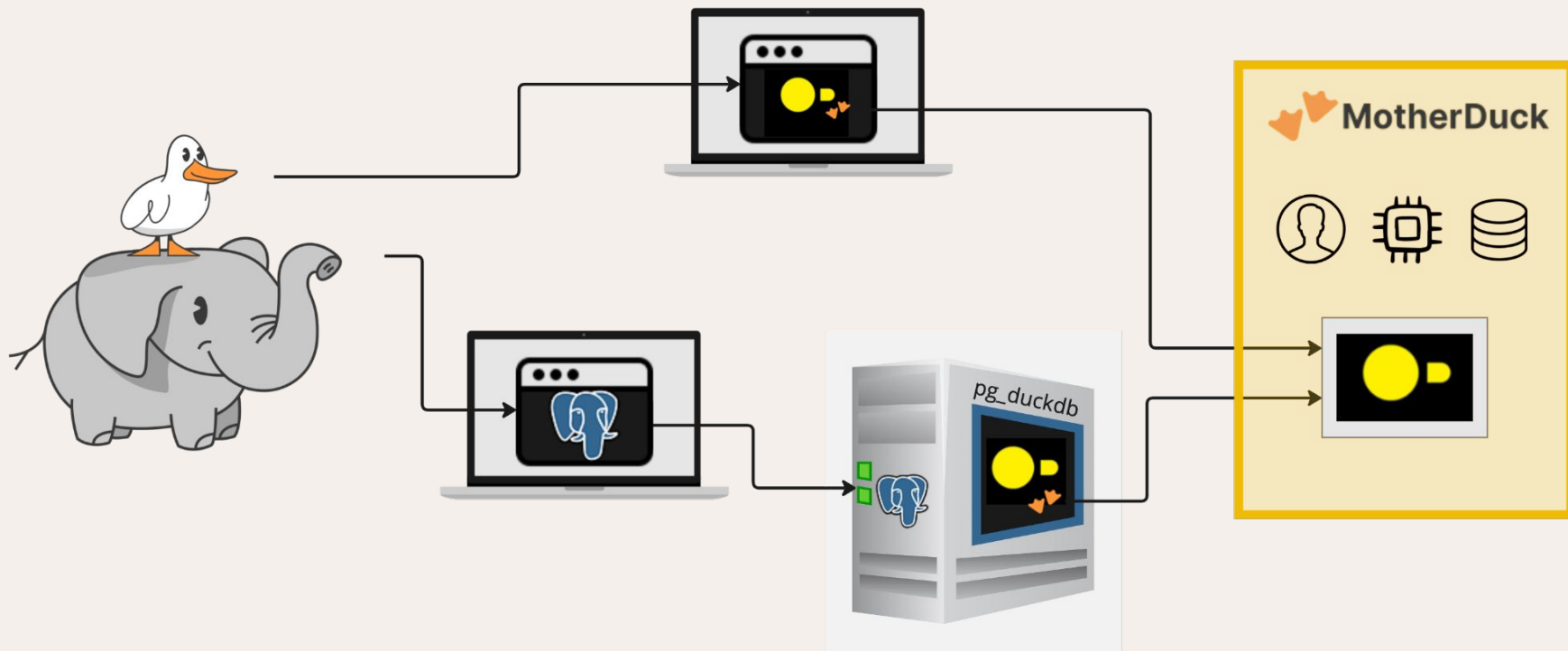
# Embedded Analytical Processing



# Embedded Analytical Processing

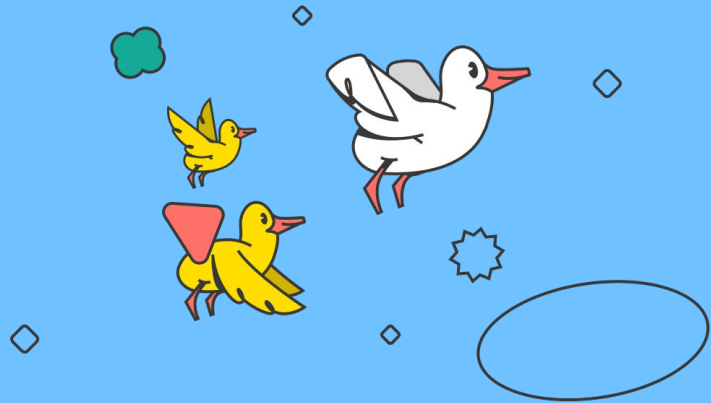


# Embedded Analytical Processing

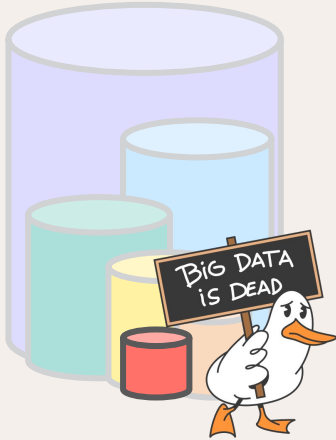




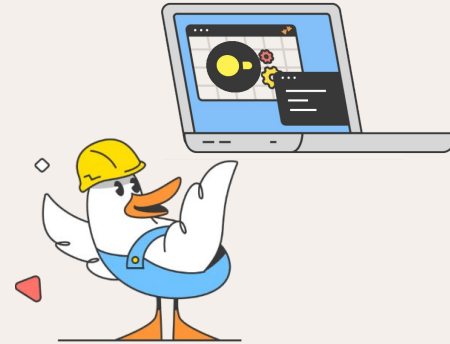
# Taking Flight



# Revisit Big Data's Assumptions

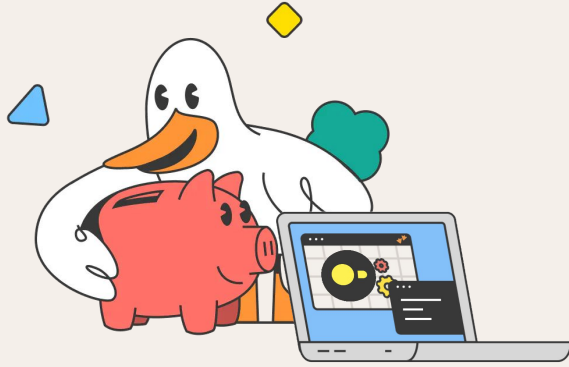


The majority of data is unused most of the time.



Laptops can do real work these days.

# Dual Execution: Local Speed, Cloud Scale



Reduce costs and increase interactivity using the compute right in front of you.



Extend to the cloud for additional compute capacity and collaboration.

# Learn More



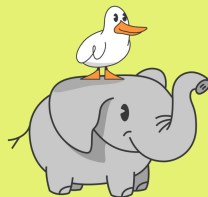
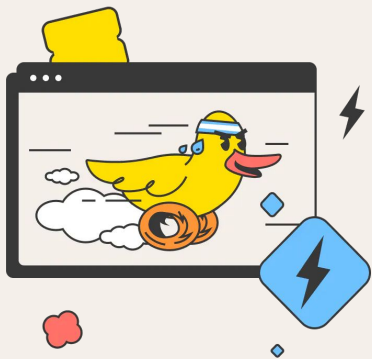
**Mother Duck**

[motherduck.com](https://motherduck.com)

**mosaic**

[uwdata.github.io/mosaic](https://uwdata.github.io/mosaic)

SEPTEMBER 23 HANDS-ON WORKSHOPS  
**SMALL DATA SF**  
SMALLDATASF.COM  
USER TALKS SEPTEMBER 24



[github.com/duckdb](https://github.com/duckdb)

**pg\_duckdb**