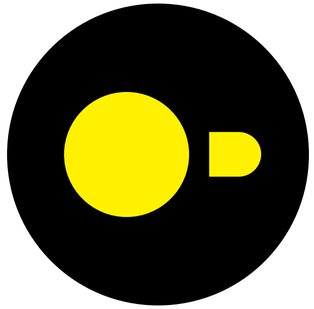


Going Beyond Two Tier Data Architectures With DuckDB

Hannes Mühleisen

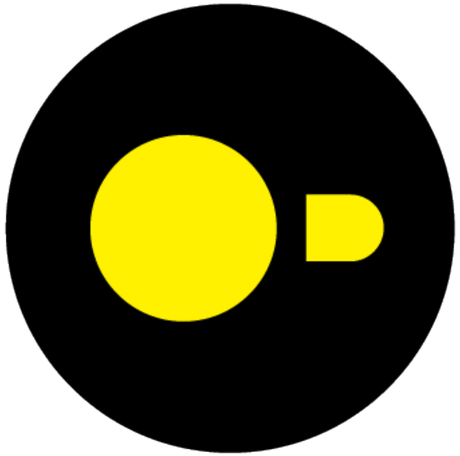
DuckDB

Co-Founder & CEO

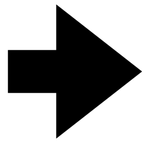
DuckDB Labs

Professor

Radboud Universiteit Nijmegen

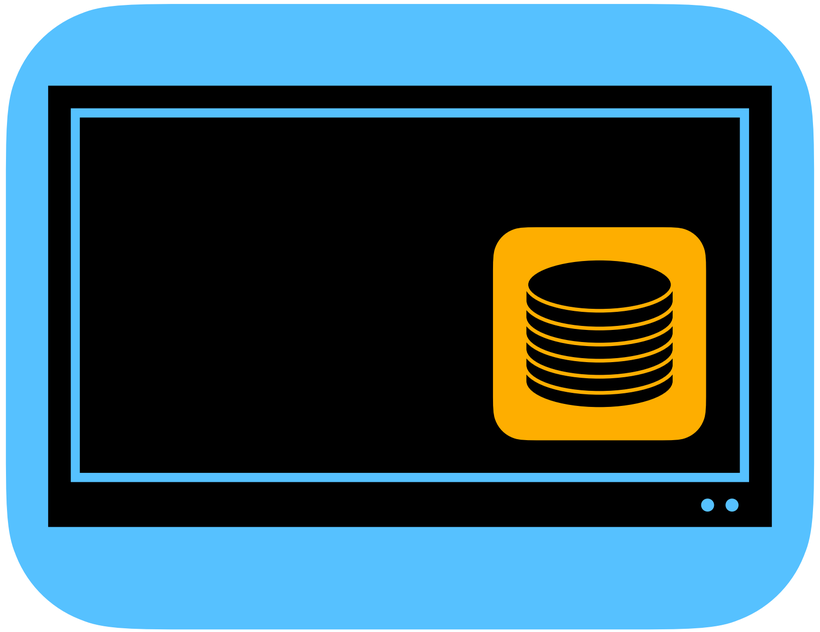


DuckDB

Data Fear  Data Confidence

User
Experience

DBMS
Engine



BATTERIES
INCLUDED

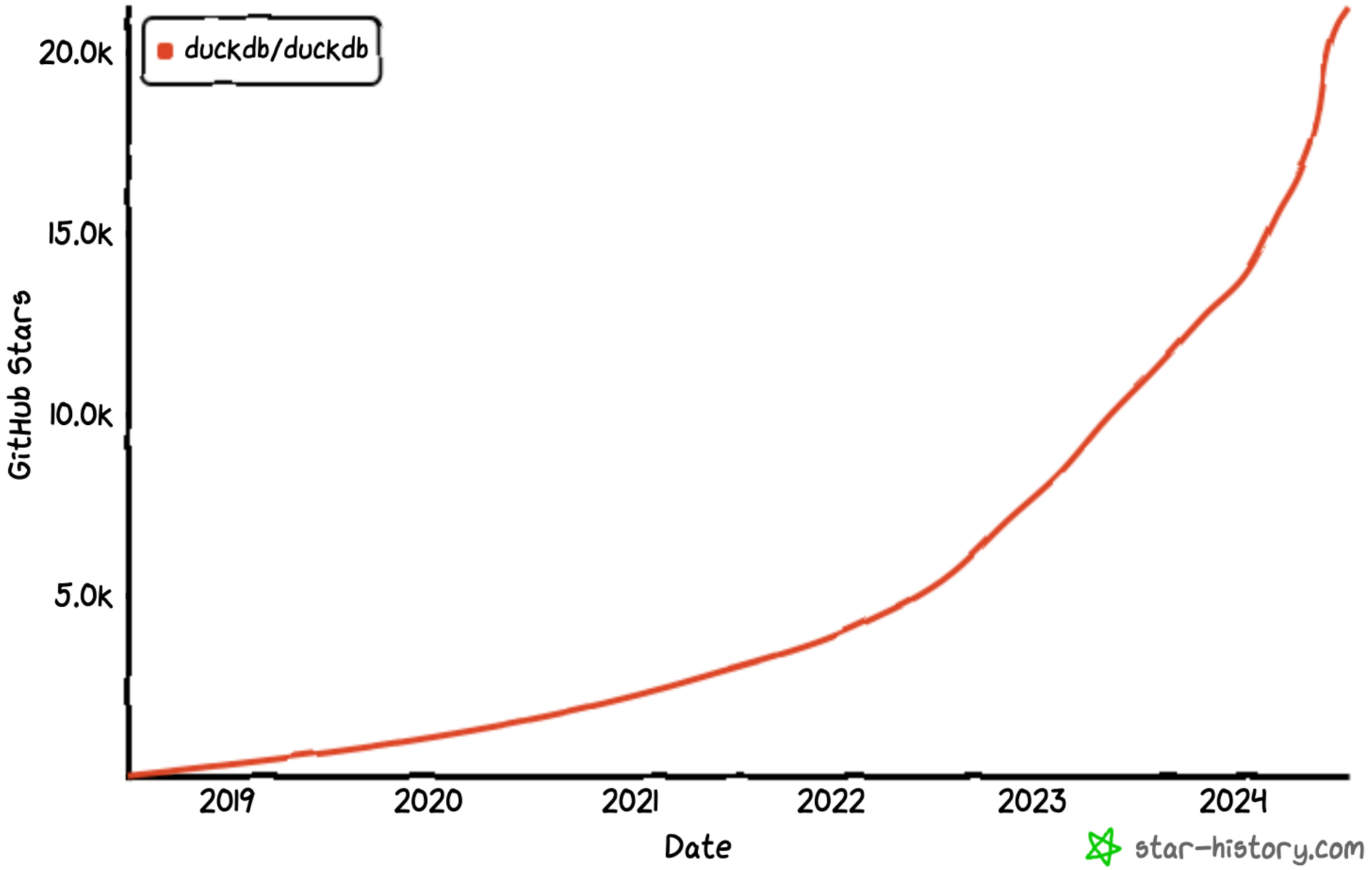
Fast

Alexander Doria
@Dorialexander

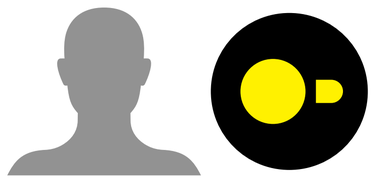
AI aside, @duckdb is probably the most magical piece of technology of recent years.

8:37 AM · Aug 14, 2024 · **5,312** Views

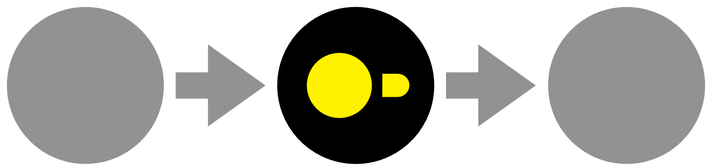
Free!


duckdb/duckdb
20.0k
15.0k
10.0k
5.0k
GitHub Stars
2019
2020
2021
2022
2023
2024
Date
star-history.com

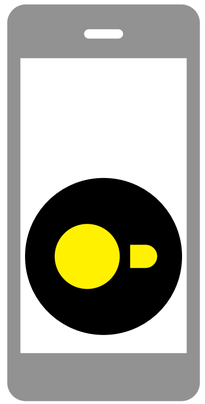
Interactive Analysis

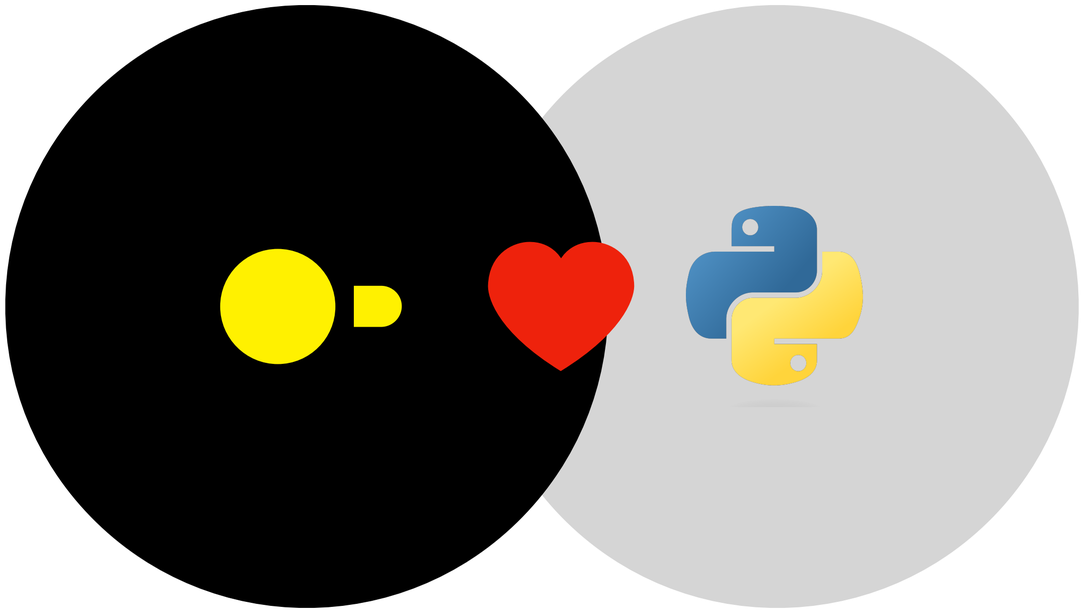


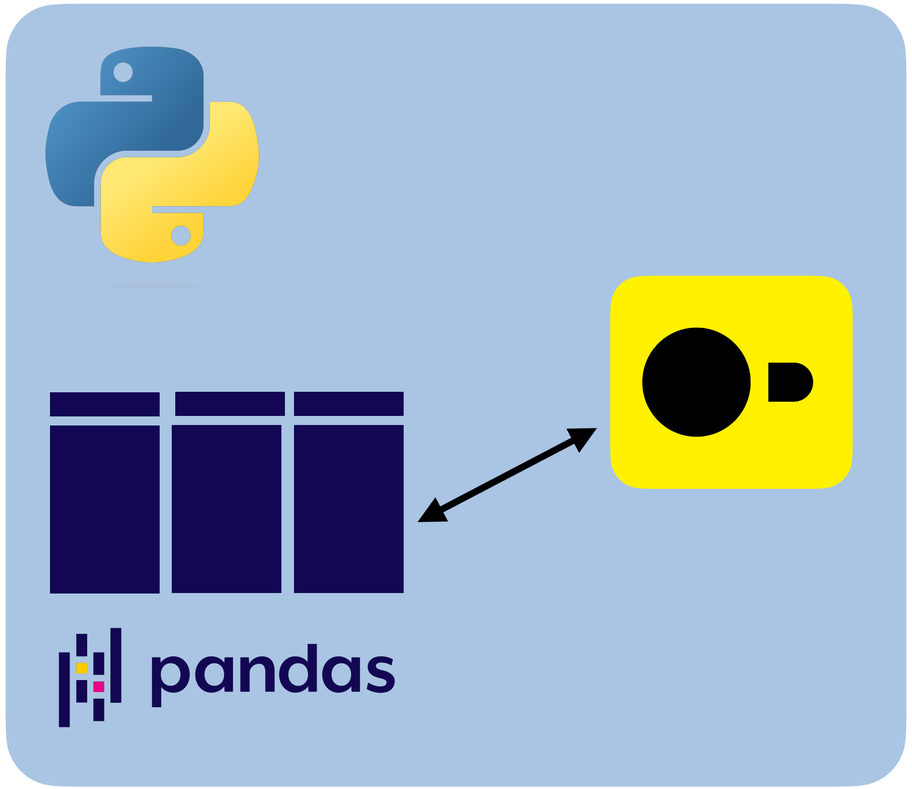
Pipeline Component

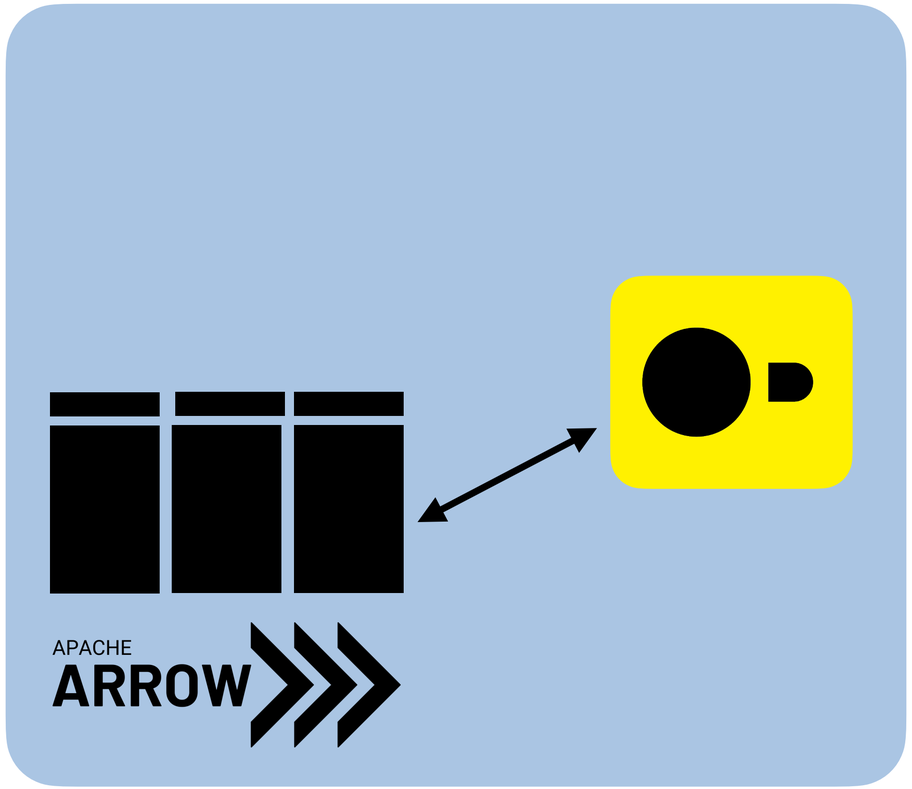


“Creative” Architecture

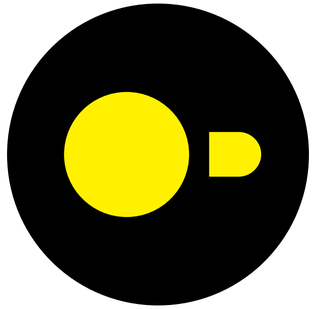


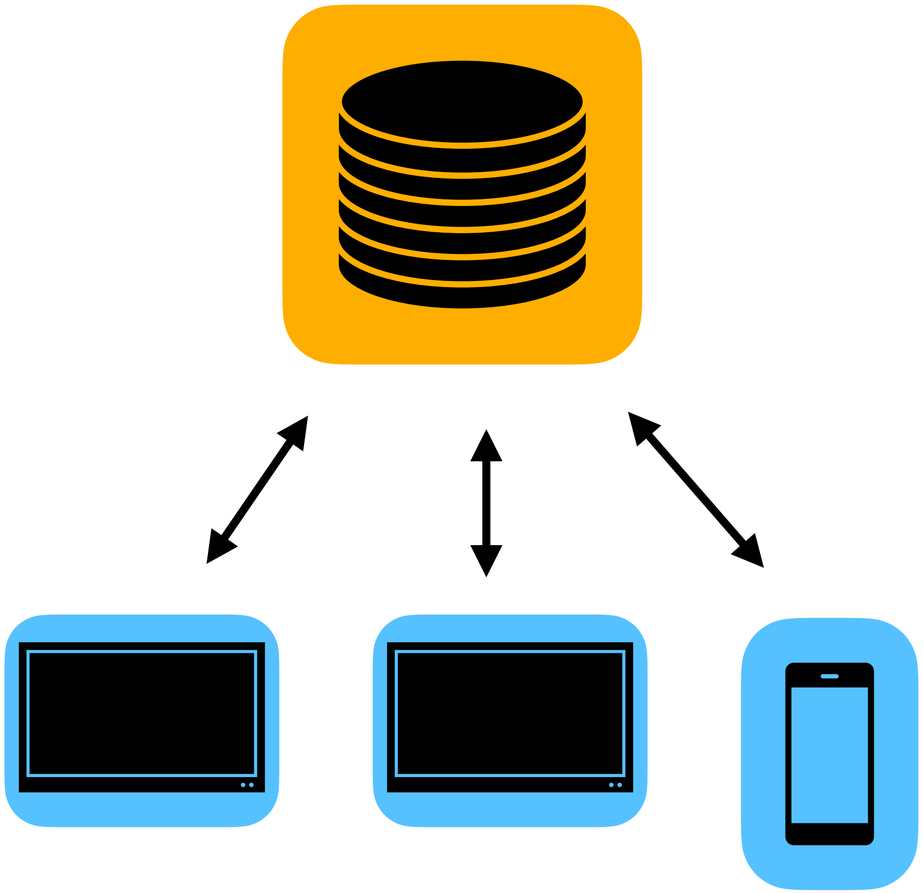


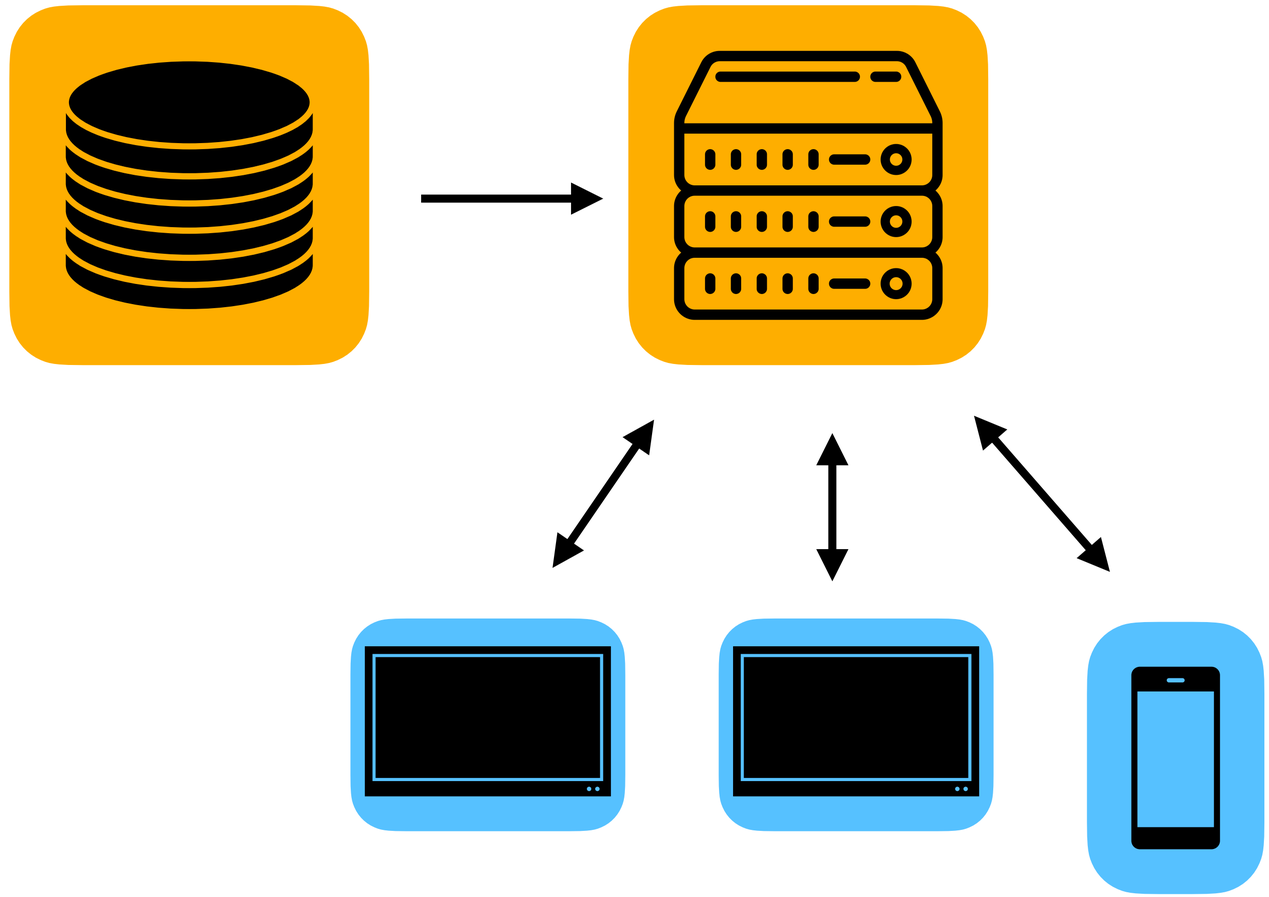

pandas
Python Process

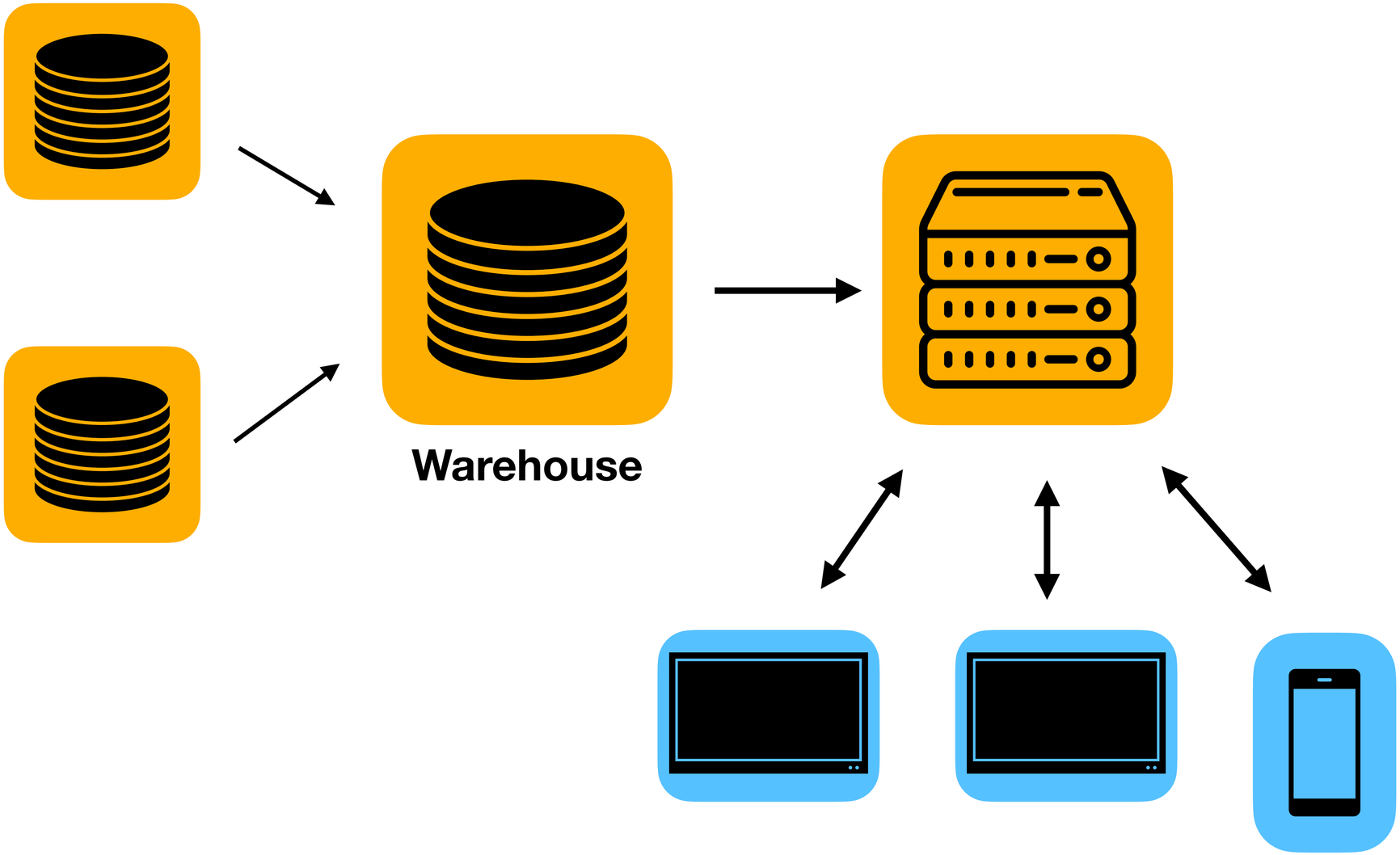


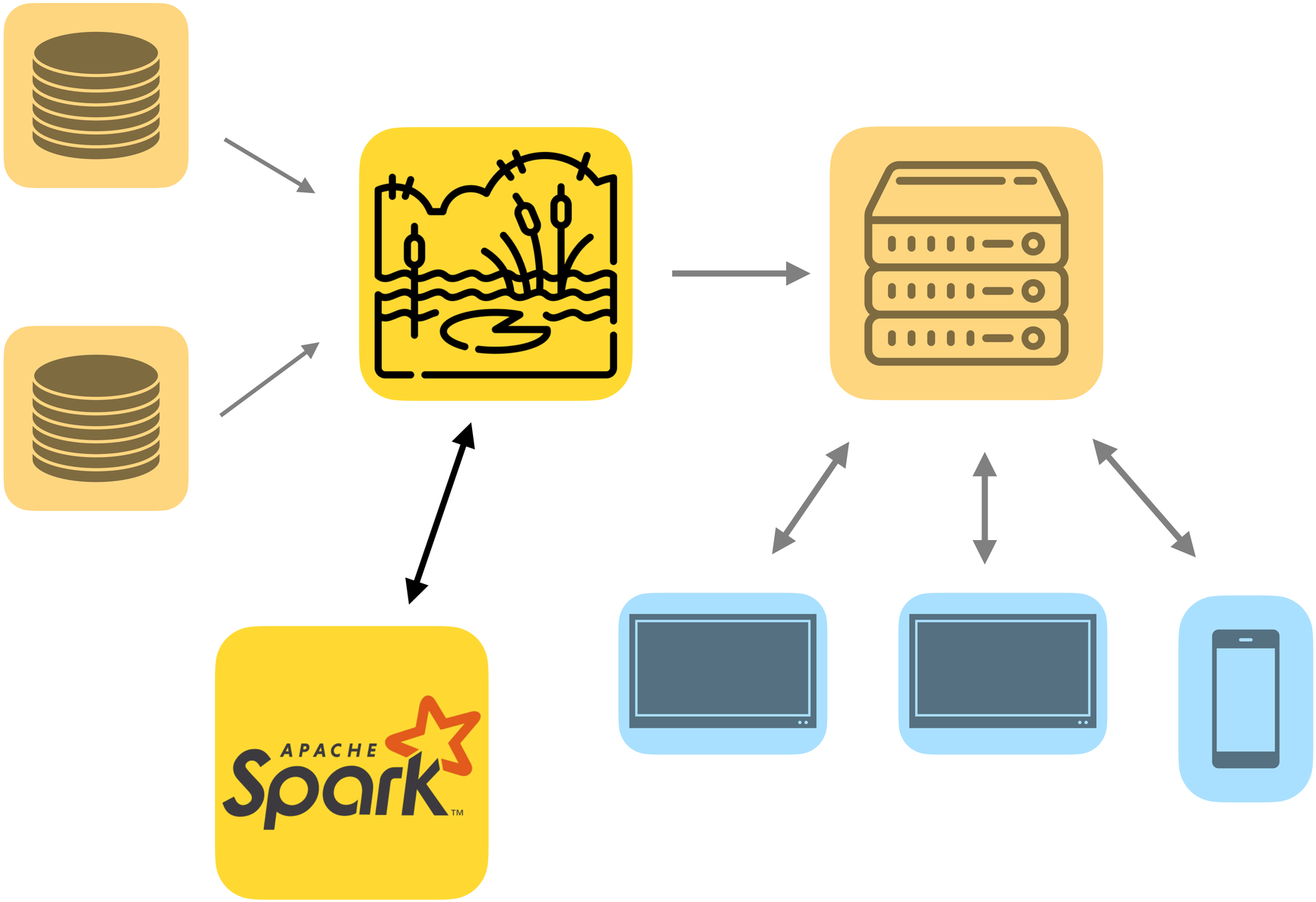
Python/R Process

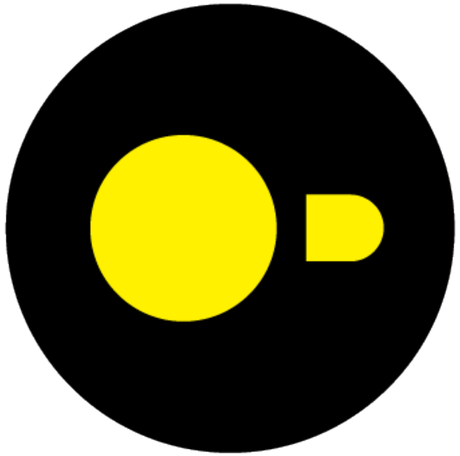




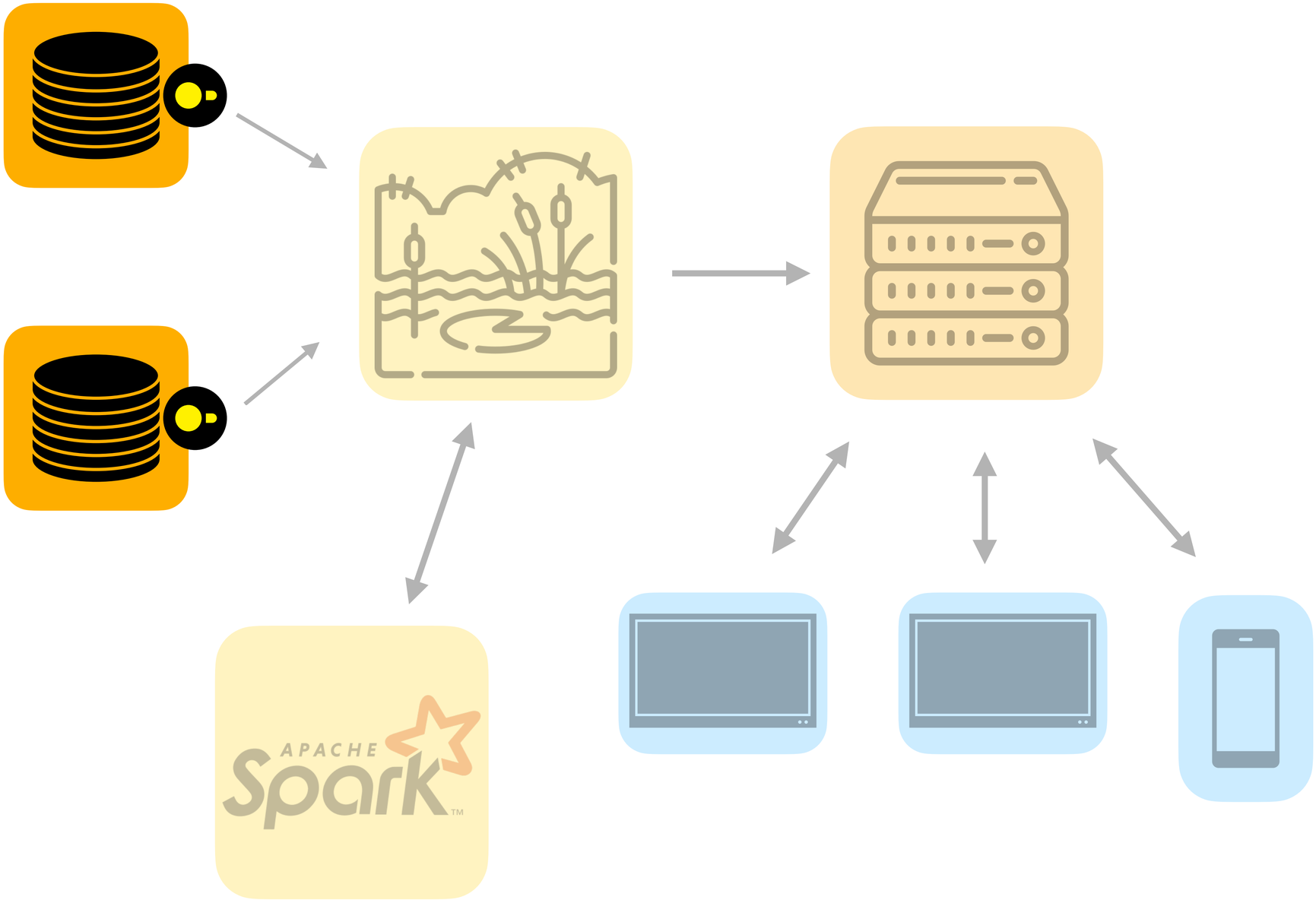



Warehouse

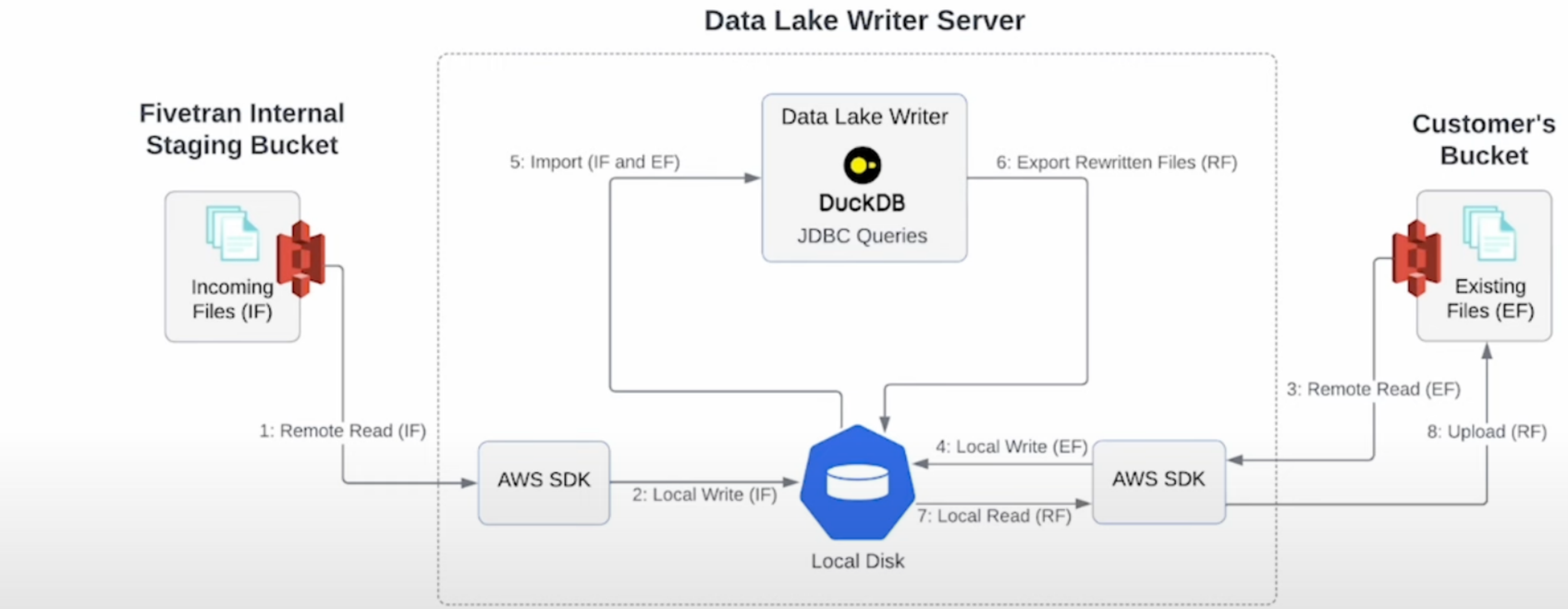

APACHE
Spark™



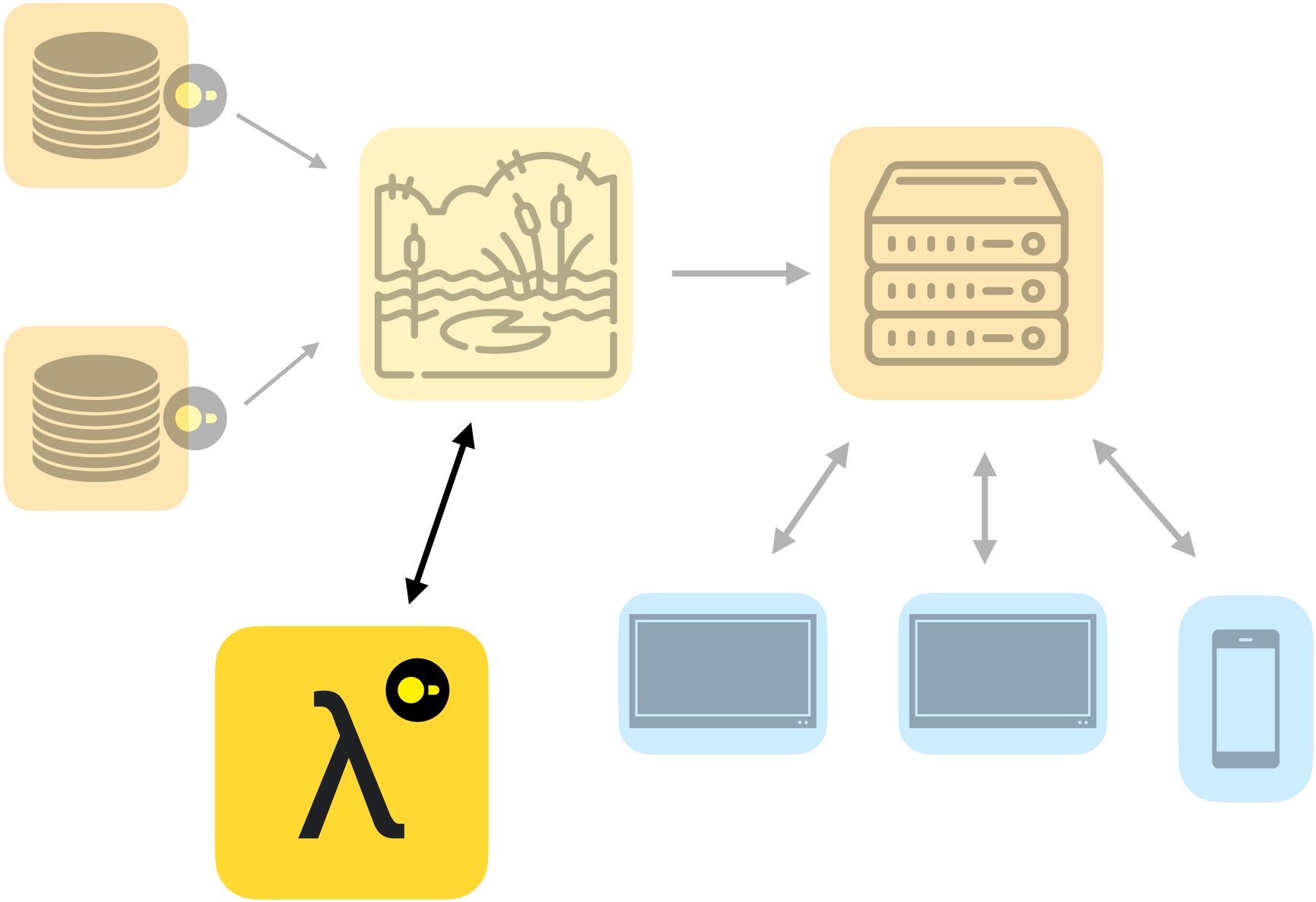
DuckDB

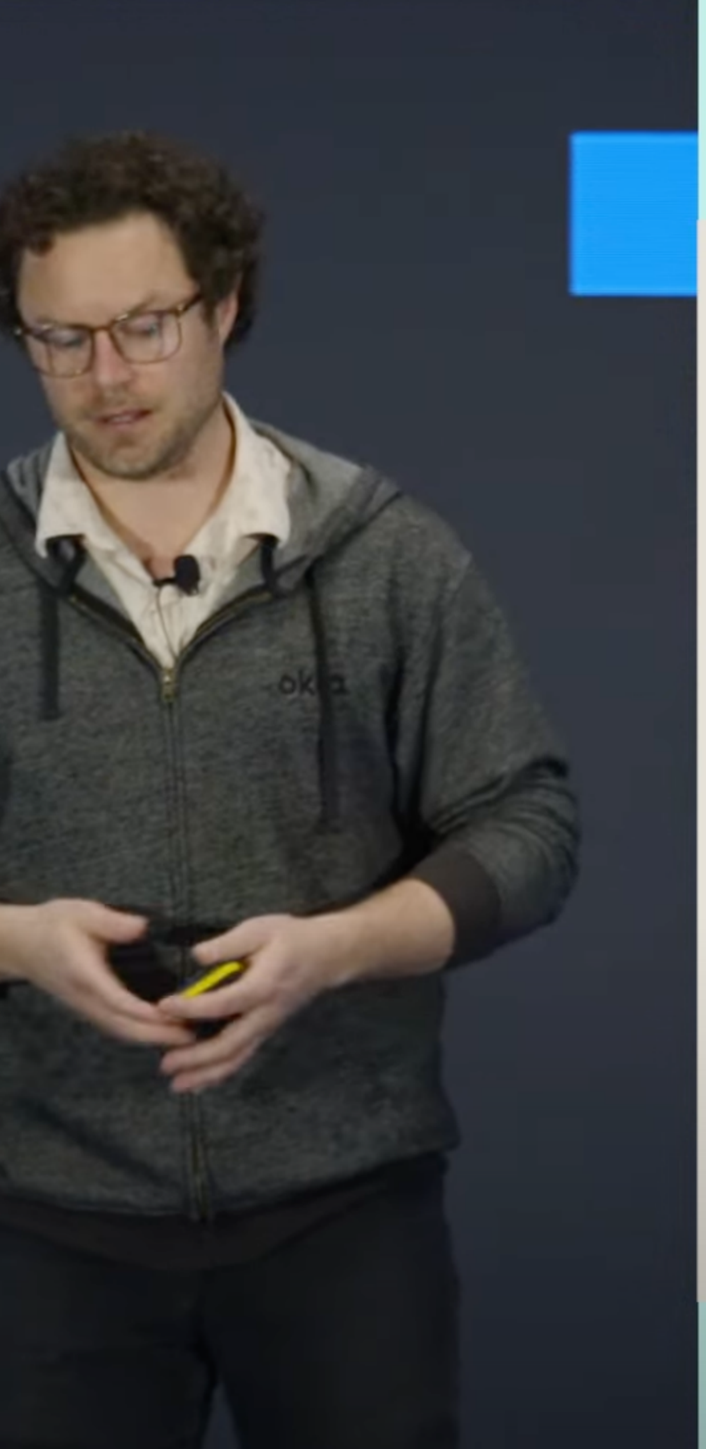

APACHE
Spark

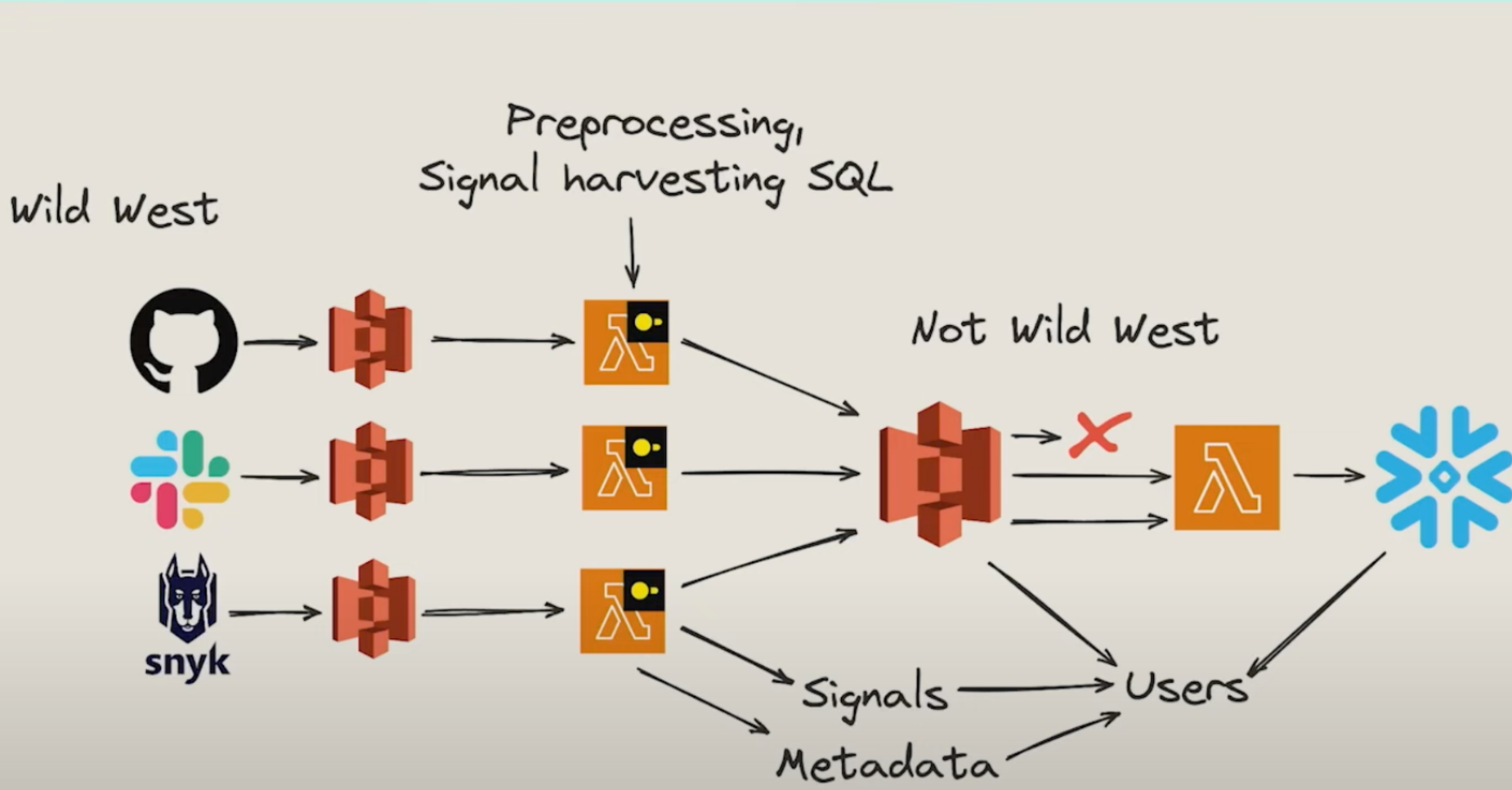
DuckDB

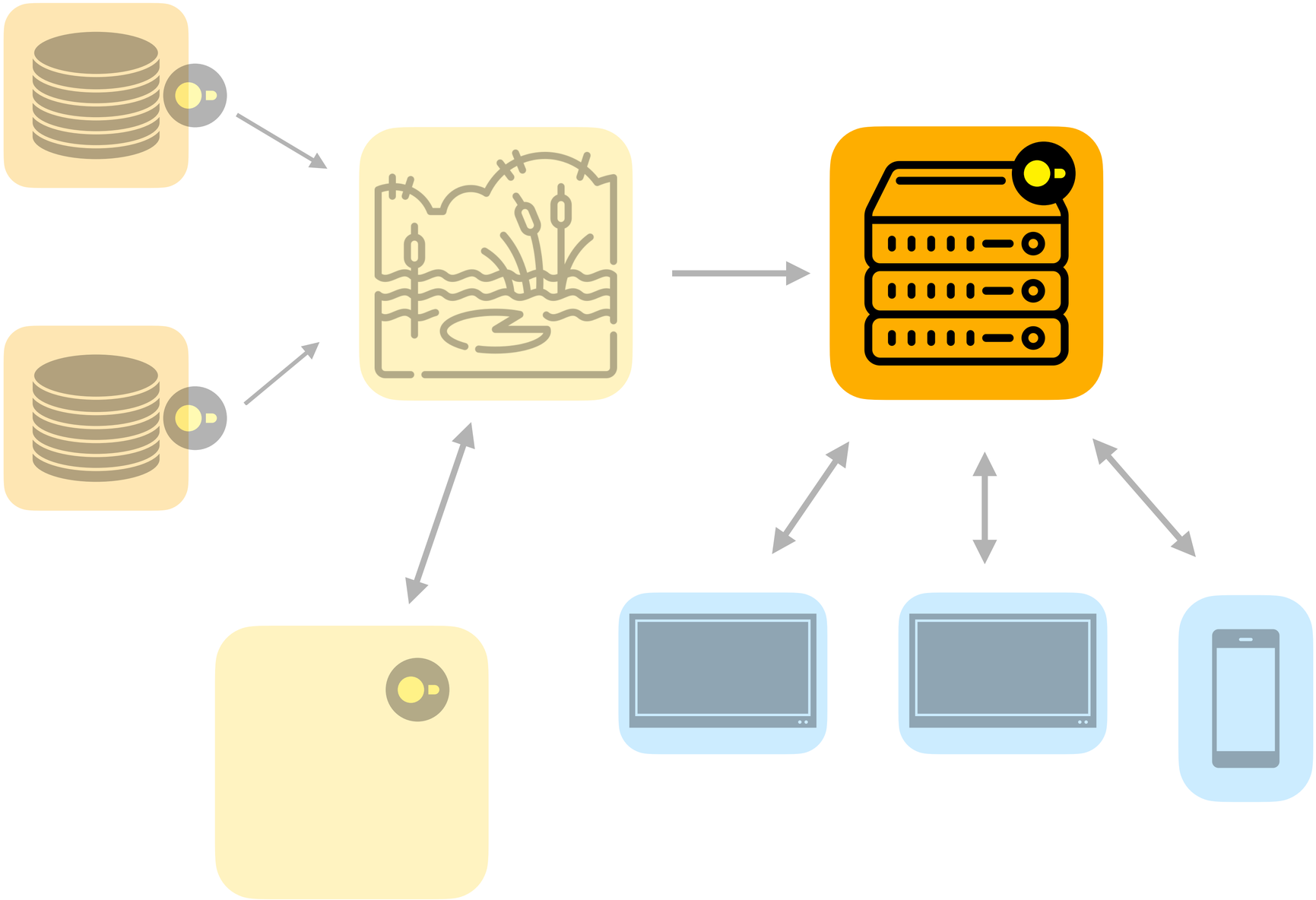


DuckCon #4 | Amsterdam






Wild West
Preprocessing,
Signal harvesting SQL
Not Wild West
snyk
Signals
Users
Metadata



r/node • 12 days ago
oubord

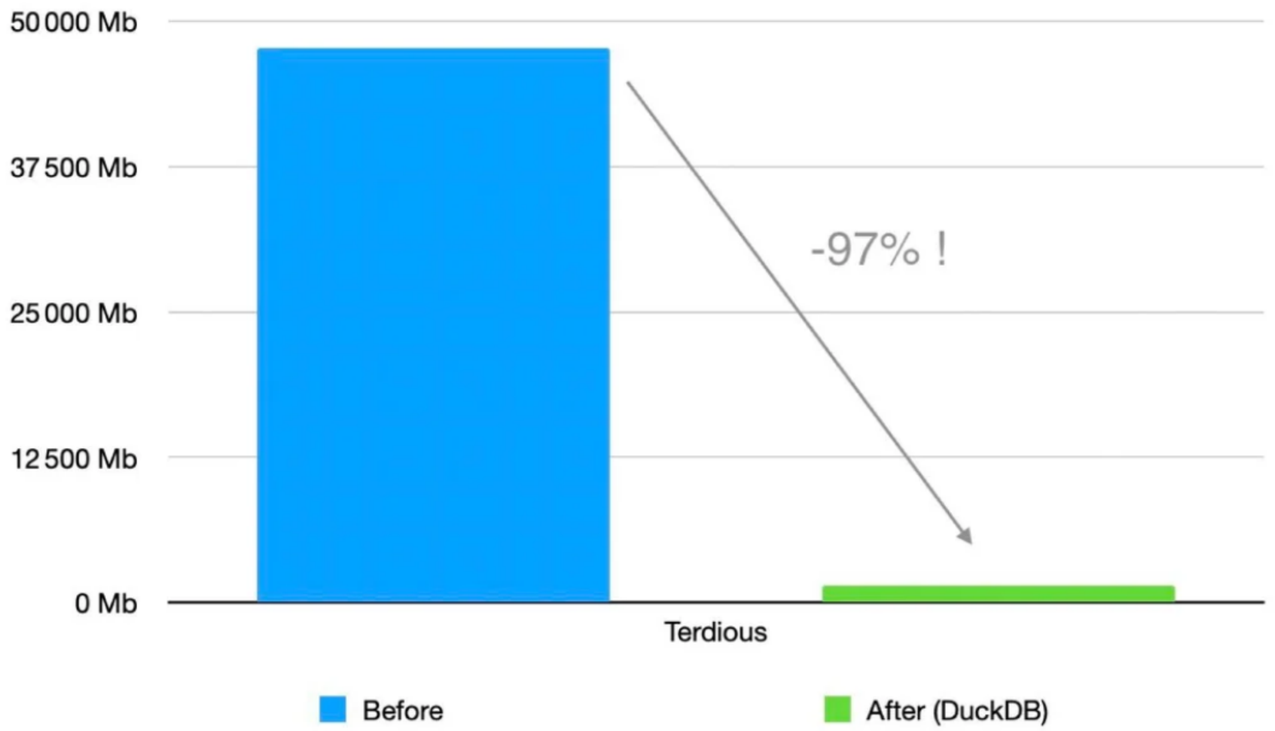
Migrated our Node.js app from SQLite to DuckDB and the performance gains are INSANE! 😲

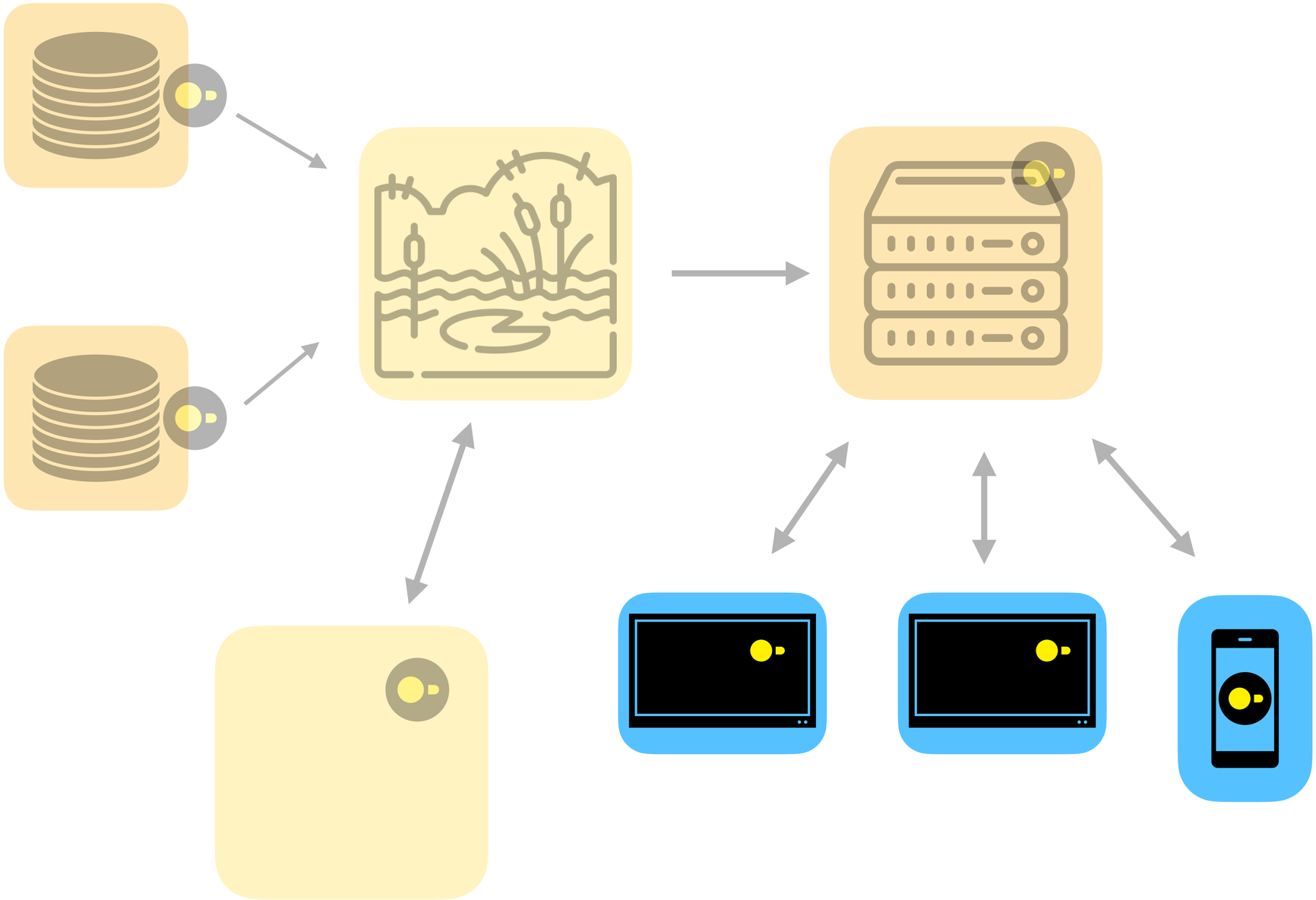
We've just made a huge change in our open-source Node.js app, Gladys Assistant, by migrating from SQLite to DuckDB for handling time-series data, and the results have been mind-blowing!

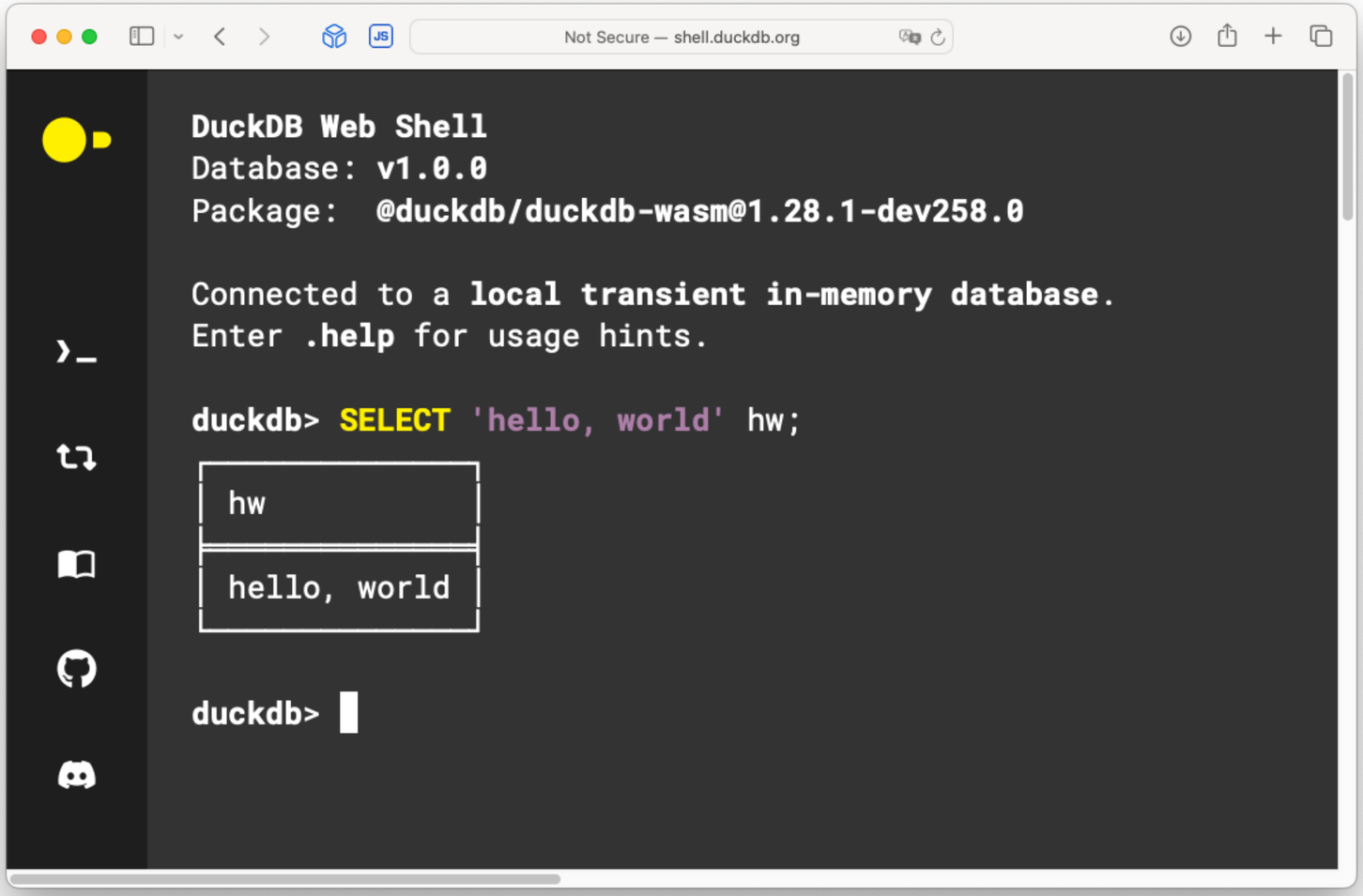
For a bit of context, Gladys Assistant is a smart home platform that connects to various sensors in your home, tracking things like temperature, electricity usage, and humidity. Over time, this generates millions of data points, which we use to display detailed charts for users.

After the switch to DuckDB, we've seen some incredible improvements:

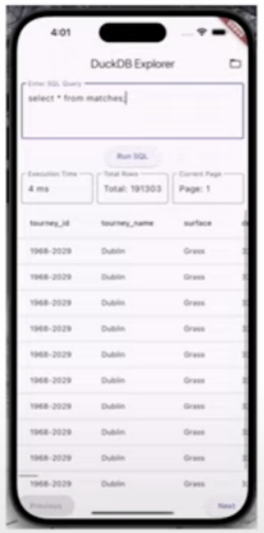
🚀 Up to a 97% reduction in database size (one user's DB went from 47 GB to just 1.5 GB).

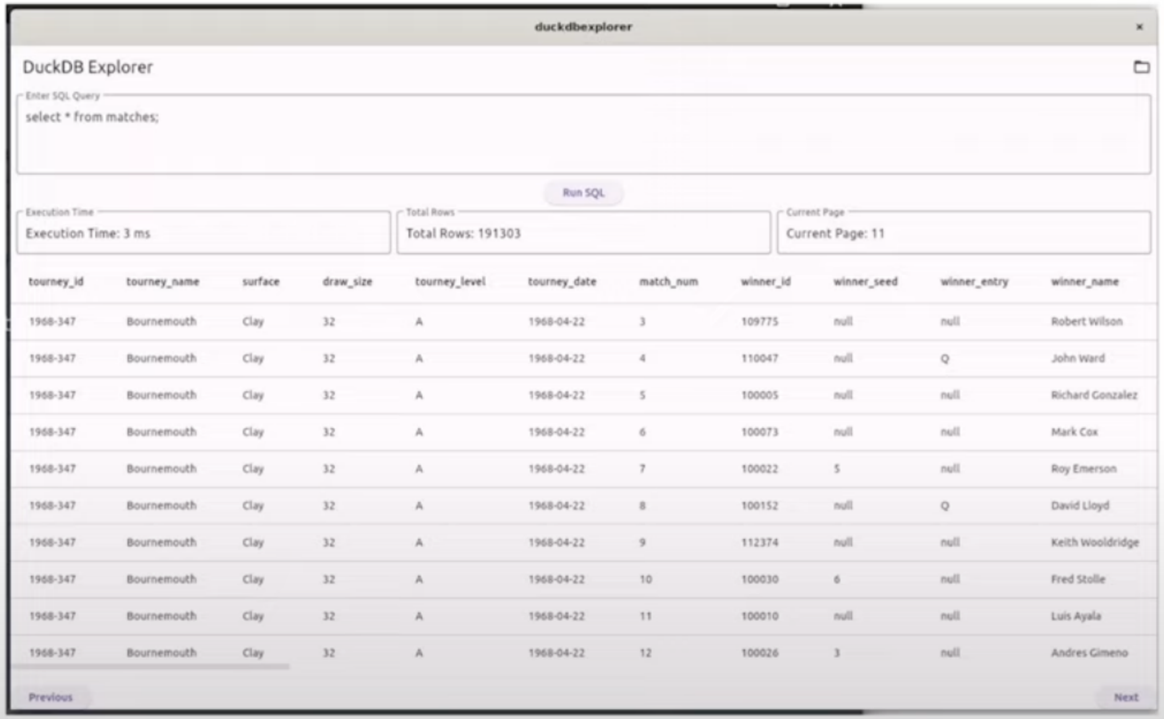





Not Secure — shell.duckdb.org
DuckDB Web Shell
Database: v1.0.0
Package: @duckdb/duckdb-wasm@1.28.1-dev258.0
Connected to a local transient in-memory database.
Enter .help for usage hints.
duckdb> SELECT 'hello, world' hw;
hw
hello, world
duckdb>

DuckDB Explorer





tourney_id	tourney_name	surface	draw_size	tourney_level	tourney_date	match_num	winner_id	winner_seed	winner_entry	winner_name
1968-347	Bournemouth	Clay	32	A	1968-04-22	3	109775	null	null	Robert Wilson
1968-347	Bournemouth	Clay	32	A	1968-04-22	4	110047	null	Q	John Ward
1968-347	Bournemouth	Clay	32	A	1968-04-22	5	100005	null	null	Richard Gonzalez
1968-347	Bournemouth	Clay	32	A	1968-04-22	6	100073	null	null	Mark Cox
1968-347	Bournemouth	Clay	32	A	1968-04-22	7	100022	5	null	Roy Emerson
1968-347	Bournemouth	Clay	32	A	1968-04-22	8	100152	null	Q	David Lloyd
1968-347	Bournemouth	Clay	32	A	1968-04-22	9	112374	null	null	Keith Wooldridge
1968-347	Bournemouth	Clay	32	A	1968-04-22	10	100030	6	null	Fred Stolle
1968-347	Bournemouth	Clay	32	A	1968-04-22	11	100010	null	null	Luis Ayala
1968-347	Bournemouth	Clay	32	A	1968-04-22	12	100026	3	null	Andres Gimeno

Datasets: SocialGrep / one-million-reddit-confessions

like 6

Modalities: Tabular | Text | Formats: csv | Languages: English | Size: 1M - 10M | Libraries: Datasets | pandas | Croissant | +1 | License: cc-by-4.0

Dataset card | Viewer | Files and versions | Community 2

Dataset Viewer

Auto-converted to Parquet | API | Embed | View in Dataset Viewer

Split (1)
train · 1M rows

Search this dataset

type string · *classes*	id string · *lengths*	subreddit.id string · *classes*	subreddit.name string · *classes*	subreddit.nsfw bool	created_ut int64
1 value	6 — 6	4 values	4 values	1 class	1.55B
post	pyw04g	2qo2a	confession	false	1,633
post	pyvzbo	2qo2a	confession	false	1,633
post	pyvxtt	2qo2a	confession	false	1,633
post	pyvwgp	2qo2a	confession	false	1,633
post	pyvt7k	2qo2a	confession	false	1,633
post	pyvsmx	2qo2a	confession	false	1,633

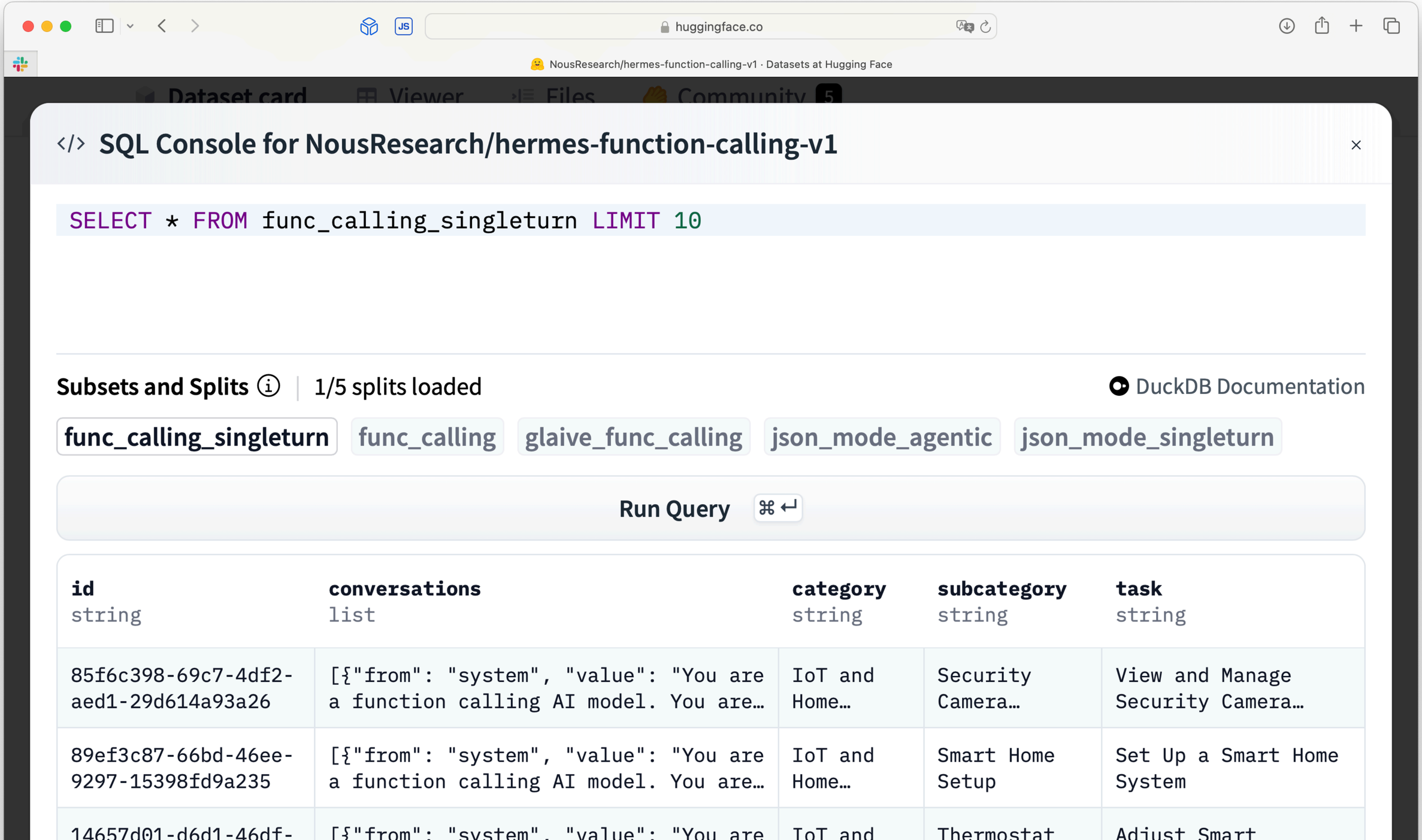
Downloads last month 16

Use this dataset | Edit dataset card

Homepage: socialgrep.com
Point of Contact: Website
Size of downloaded dataset files: 779 MB
Size of the auto-converted Parquet files: 439 MB
Number of rows: 1,000,000

Models trained or fine-tuned on SocialGrep/one-…

or4cl3ai/SoundSlayerAI
Text-to-Speech • Updated Sep 30, 2023 • 7 • 14


huggingface.co
NousResearch/hermes-function-calling-v1 · Datasets at Hugging Face
Dataset card
Viewer
Files
Community
5
SQL Console for NousResearch/hermes-function-calling-v1
SELECT * FROM func_calling_singleturn LIMIT 10
Subsets and Splits
1/5 splits loaded
DuckDB Documentation
func_calling_singleturn
func_calling
glaive_func_calling
json_mode_agentic
json_mode_singleturn
Run Query
⌘ ↵
id
string
conversations
list
category
string
subcategory
string
task
string
85f6c398-69c7-4df2-aed1-29d614a93a26
[{"from": "system", "value": "You are a function calling AI model. You are…
IoT and Home…
Security Camera…
View and Manage Security Camera…
89ef3c87-66bd-46ee-9297-15398fd9a235
[{"from": "system", "value": "You are a function calling AI model. You are…
IoT and Home…
Smart Home Setup
Set Up a Smart Home System

loader-duckdb

duckcon-cheat.md | duckcon.md

src > duckcon.md > ## Hello DuckCon

```
## Hello DuckCon

```sql id=oneYear
SELECT * FROM languages WHERE TIME_PERIOD = ${year}
```

```js
const year = view(Inputs.range([2015, 2020], {step: 1}))
```

```js
Plot.plot({
  x: {axis: "top", grid: true},
  color: {legend: true, unknown: "grey"},
  marks: [
    Plot.barX(oneYear, {
      x: "OBS_VALUE",
      y: "geo",
      fill: "language",
      order: "-sum",
```

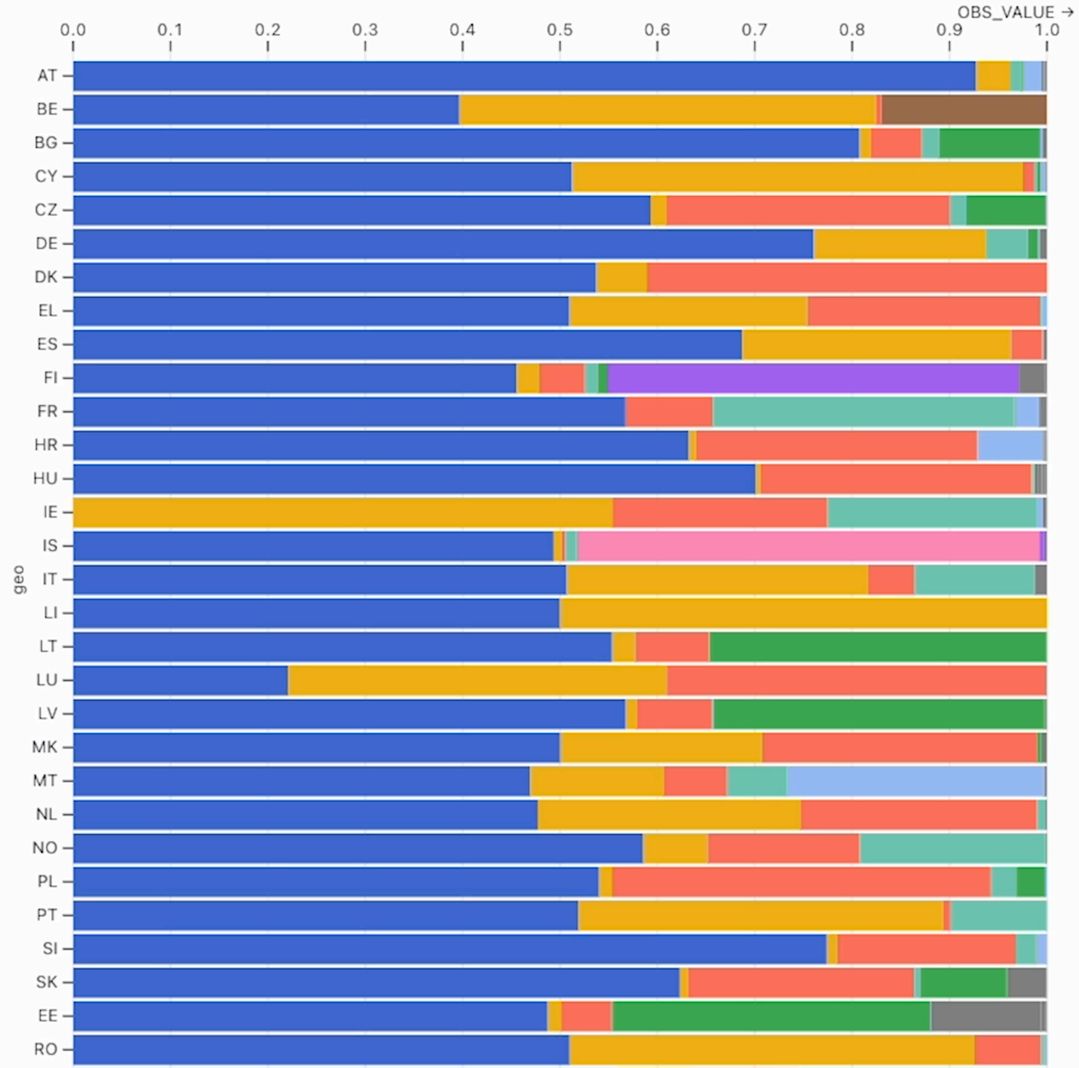
Launchpad ⊗ 0 ⚠ 0 0 Live Share

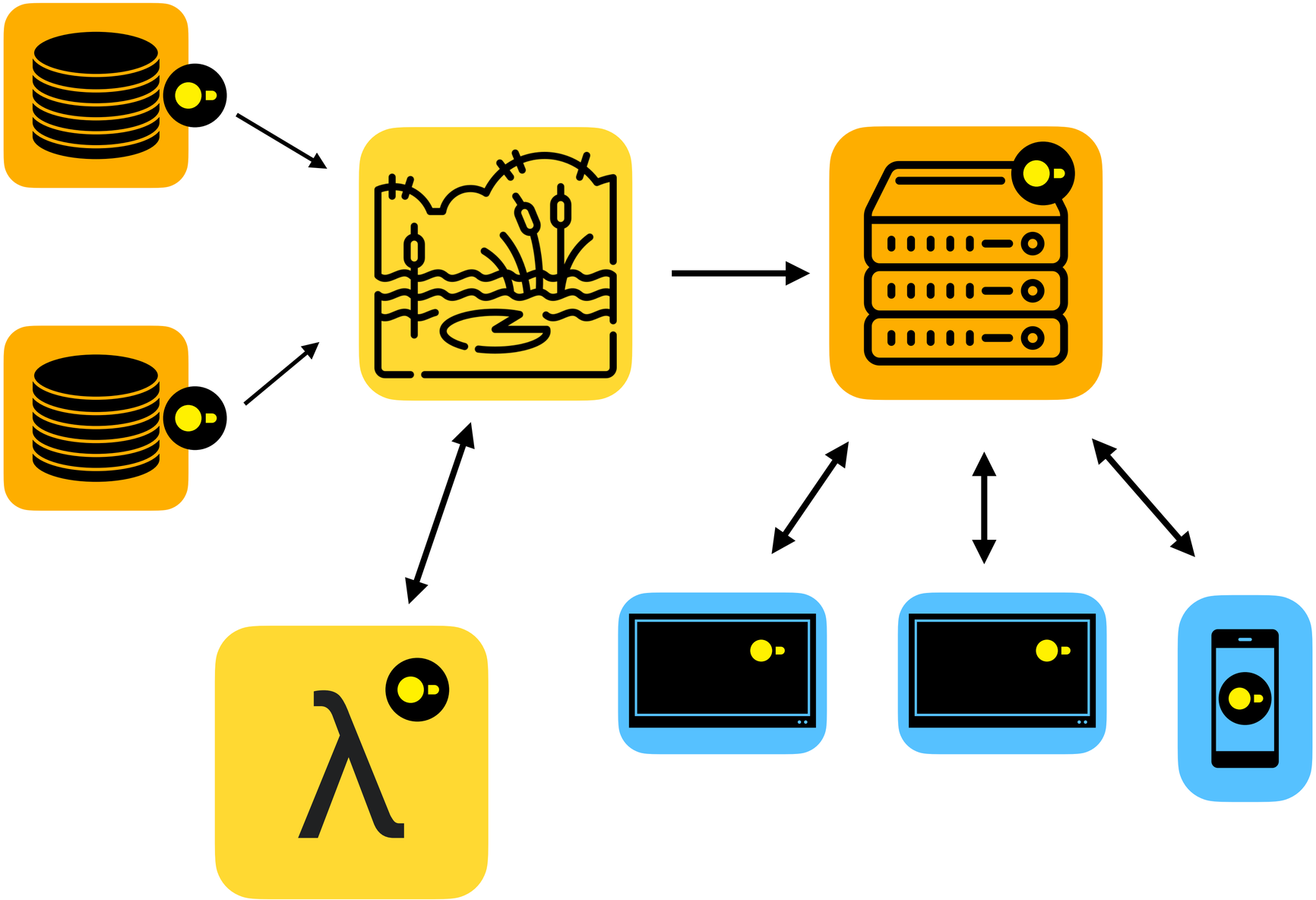
127.0.0.1:3000/duckcon

Hello DuckCon

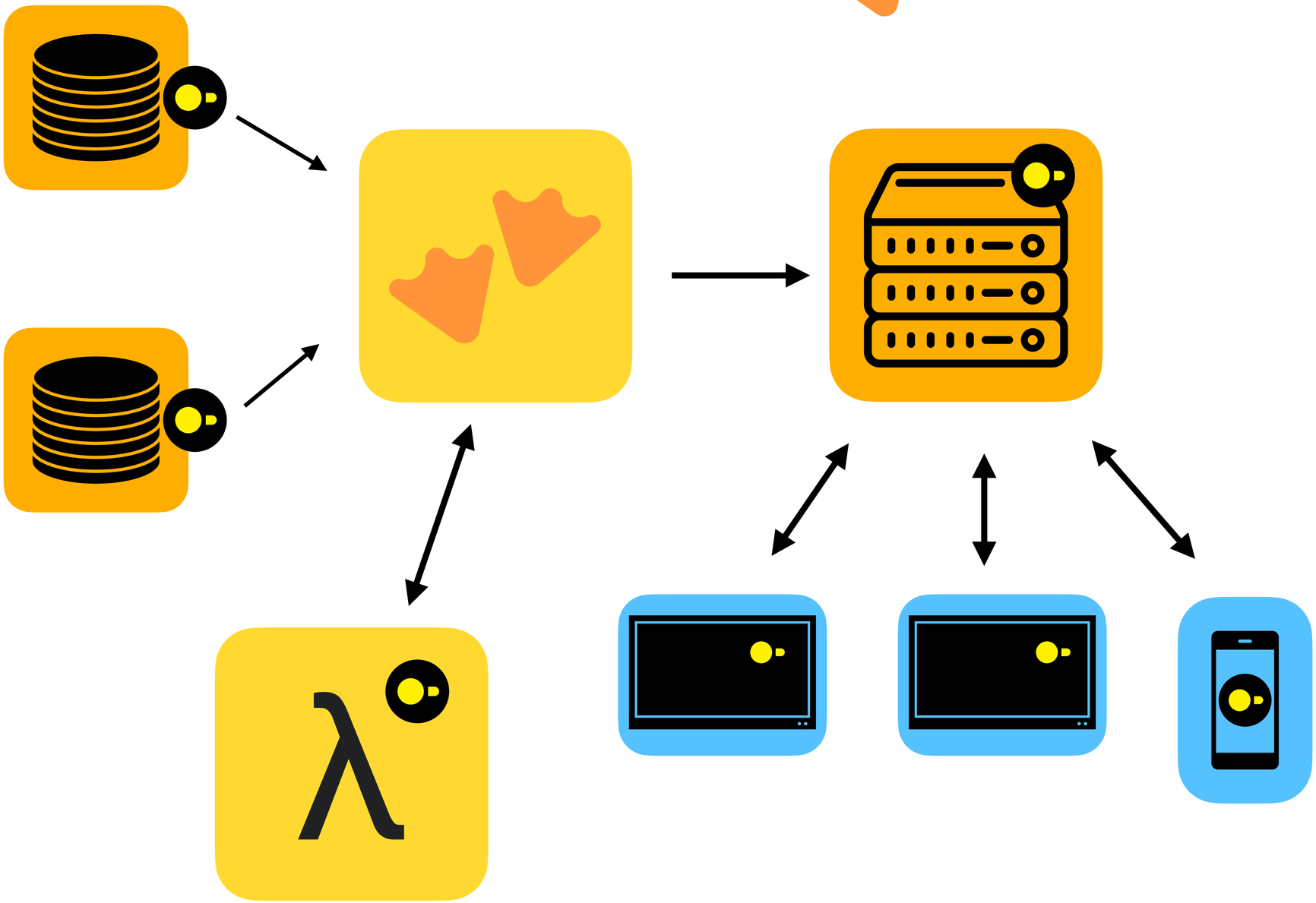
2019

ENG FRE GER SPA RUS DAN SWE ITA DUT

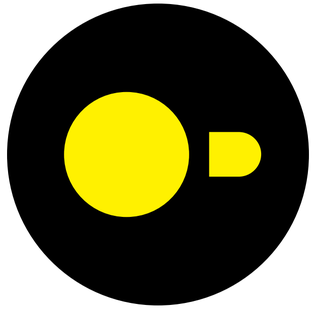

```



MotherDuck



Going Beyond Two Tier Data Architectures With DuckDB

- **Liberate** data processing
  - No more centralized data engines
  - DuckDB everywhere!

hannes.muehleisen.org
@hfmuehleisen