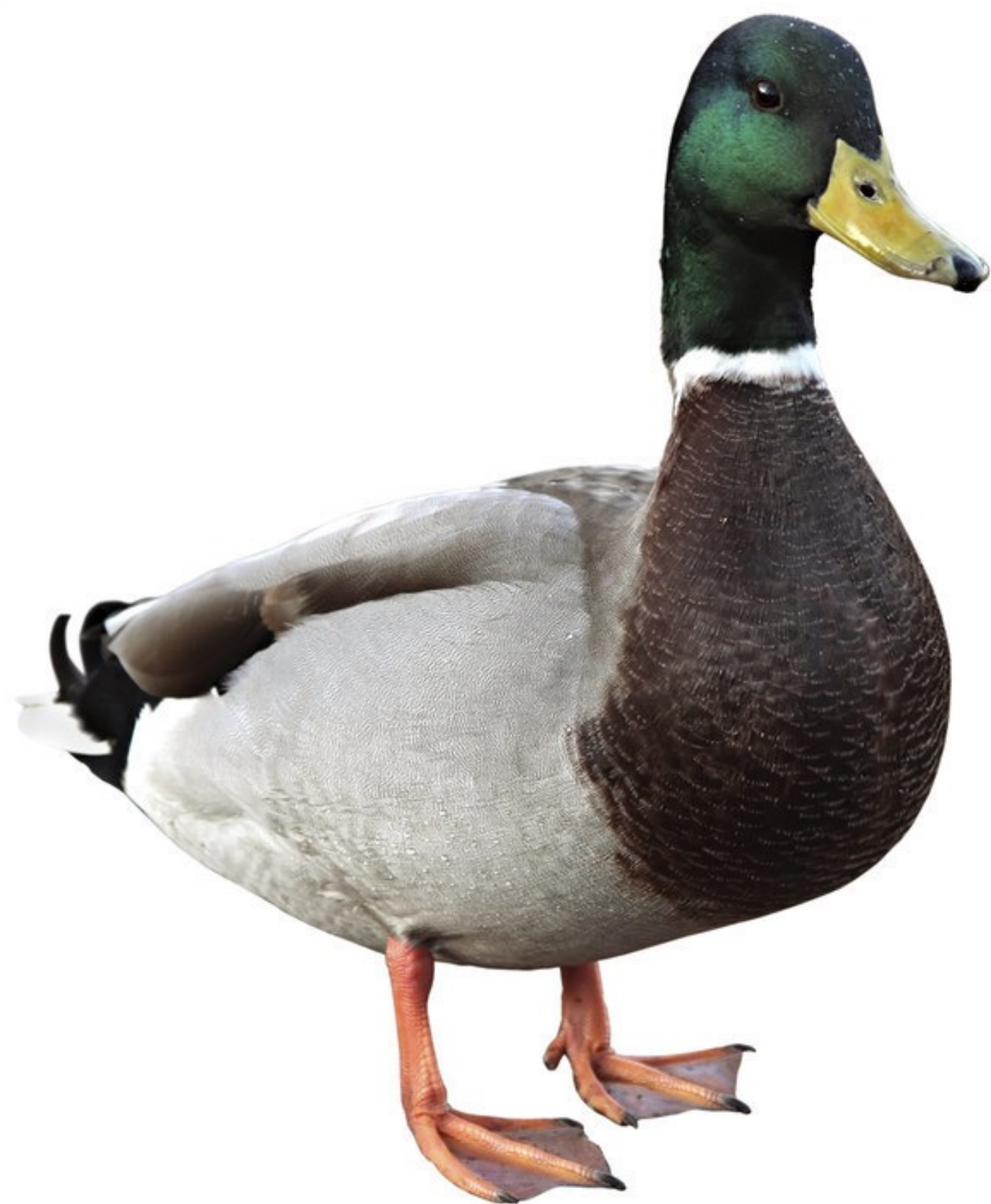
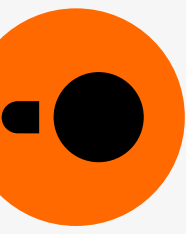




Welcome!



- 1:30 PM - Session 1
 - Overview and Latest Developments
 - MotherDuck: Taking flight with interactive analytics (Frances)
 - Lightning Talks (Part 1, Jaan & Andy)
 - Break

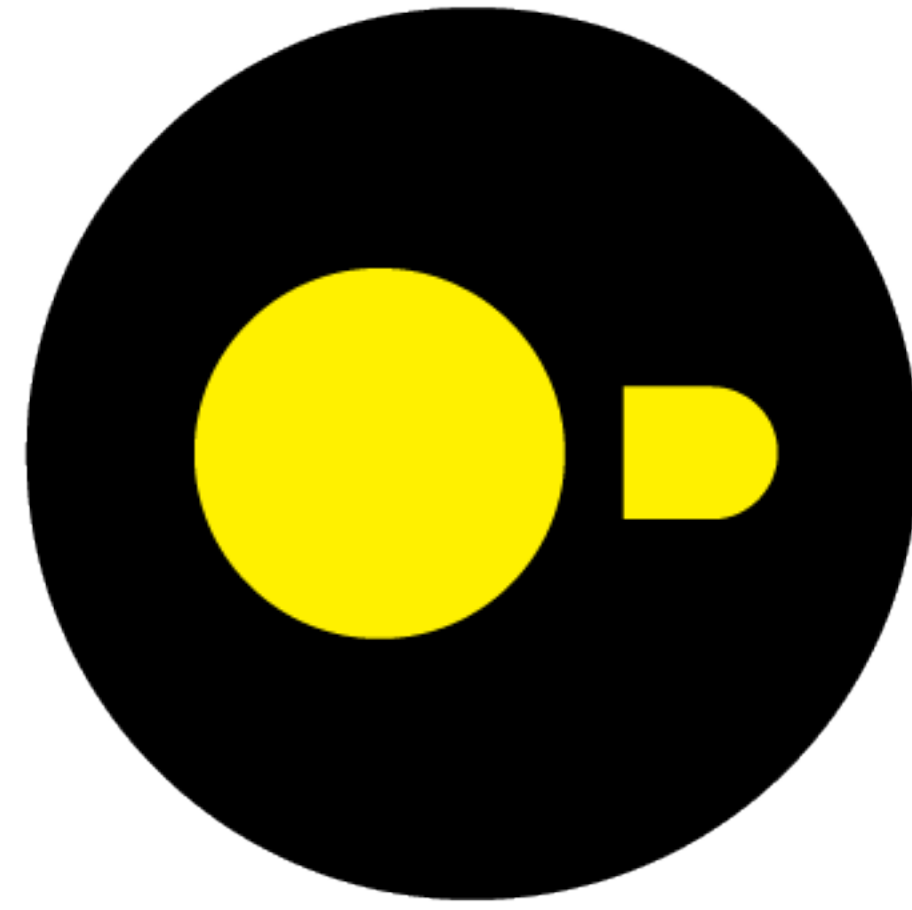


- 3:30 PM - Session 2
 - Performant data apps in the browser with DuckDB (Robert)
 - pg_duckdb, DuckDB-Powered Postgres (Joe)
 - Lightning Talks (Part 2, Miguel, Brian, Richard, Edward, Junaid)
- 5 PM - Drinks & Snacks sponsored by MotherDuck

DuckDB Overview & Latest Developments

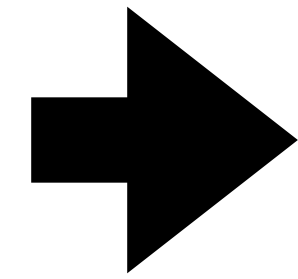


- **Where we are**
 - Adoption
- Where we're going
 - Sneak Peek 1.1.0
- Q&A



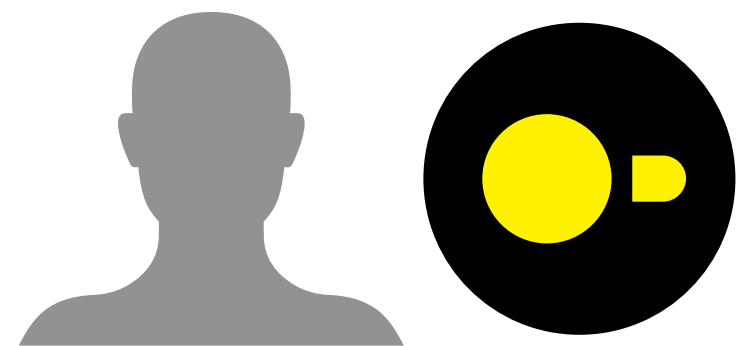
DuckDB

Data **Fear**

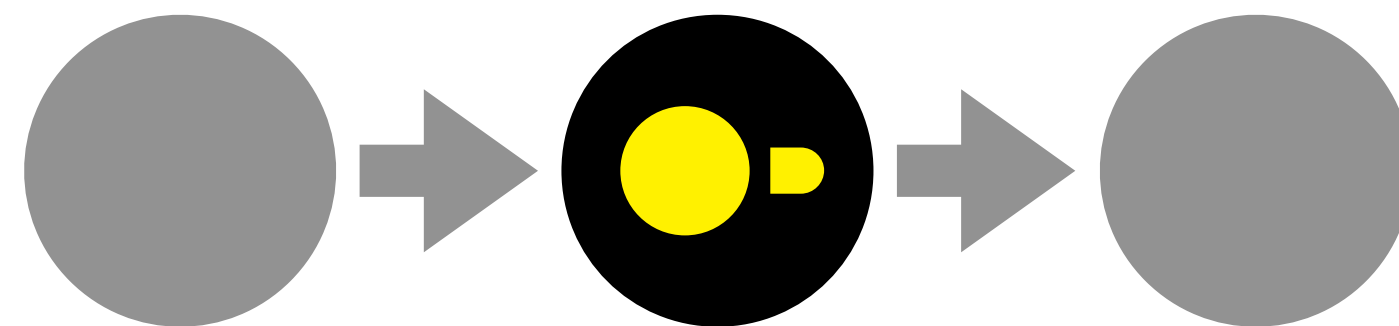


Data **Confidence**

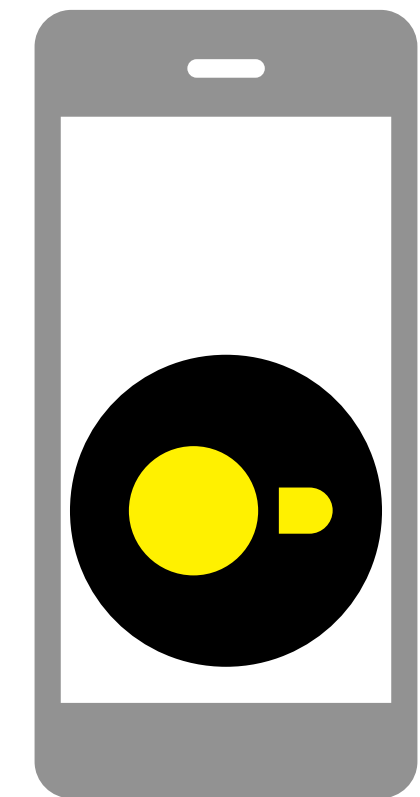
Interactive Analysis



Pipeline Component



“Creative” Architecture



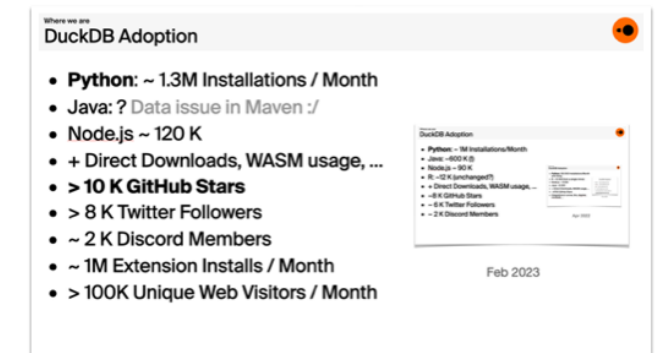


DuckDB Adoption - August 2024

- > **6 M** Downloads / Month
- > **17 M** Extension Installs / Month
- > **1 M** Unique Web Visitors / Month
- > **21 K** GitHub Stars
- > **14 K** Twitter Followers
- > **23 K** LinkedIn Followers
- > **5 K** Discord Members
- ^ DB-Engines Ranking
Overall: 62, Relational: 35

DuckDB Adoption - January 2024

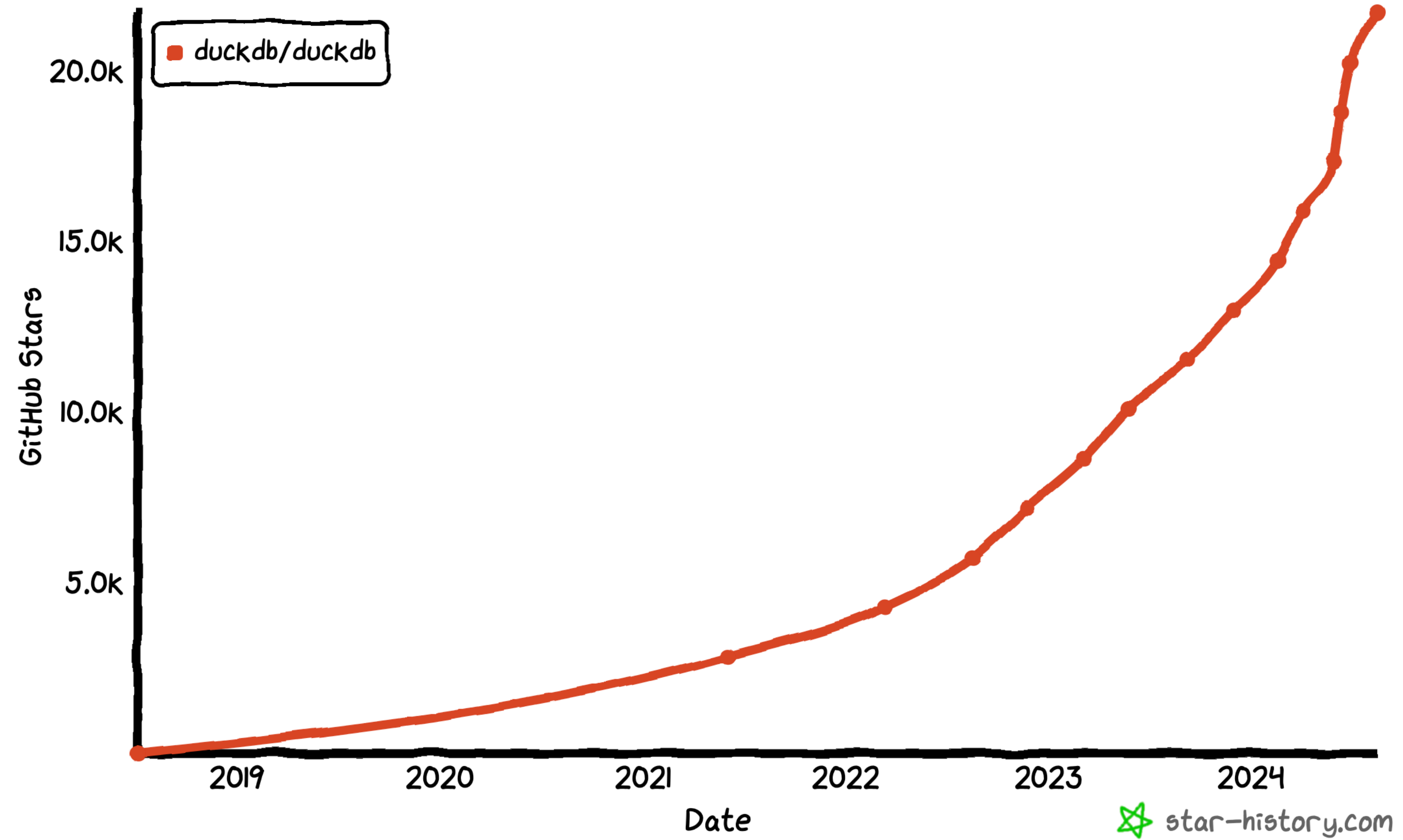
- > 2 M Downloads / Month
- > 5 M Extension Installs / Month
- > 500K Unique Web Visitors / Month
- > 14 K GitHub Stars
- > 11 K Twitter Followers
- > 6 K LinkedIn Followers
- > 4 K Discord Members
- ^ DB-Engines Ranking
Overall: 82, Relational: 45

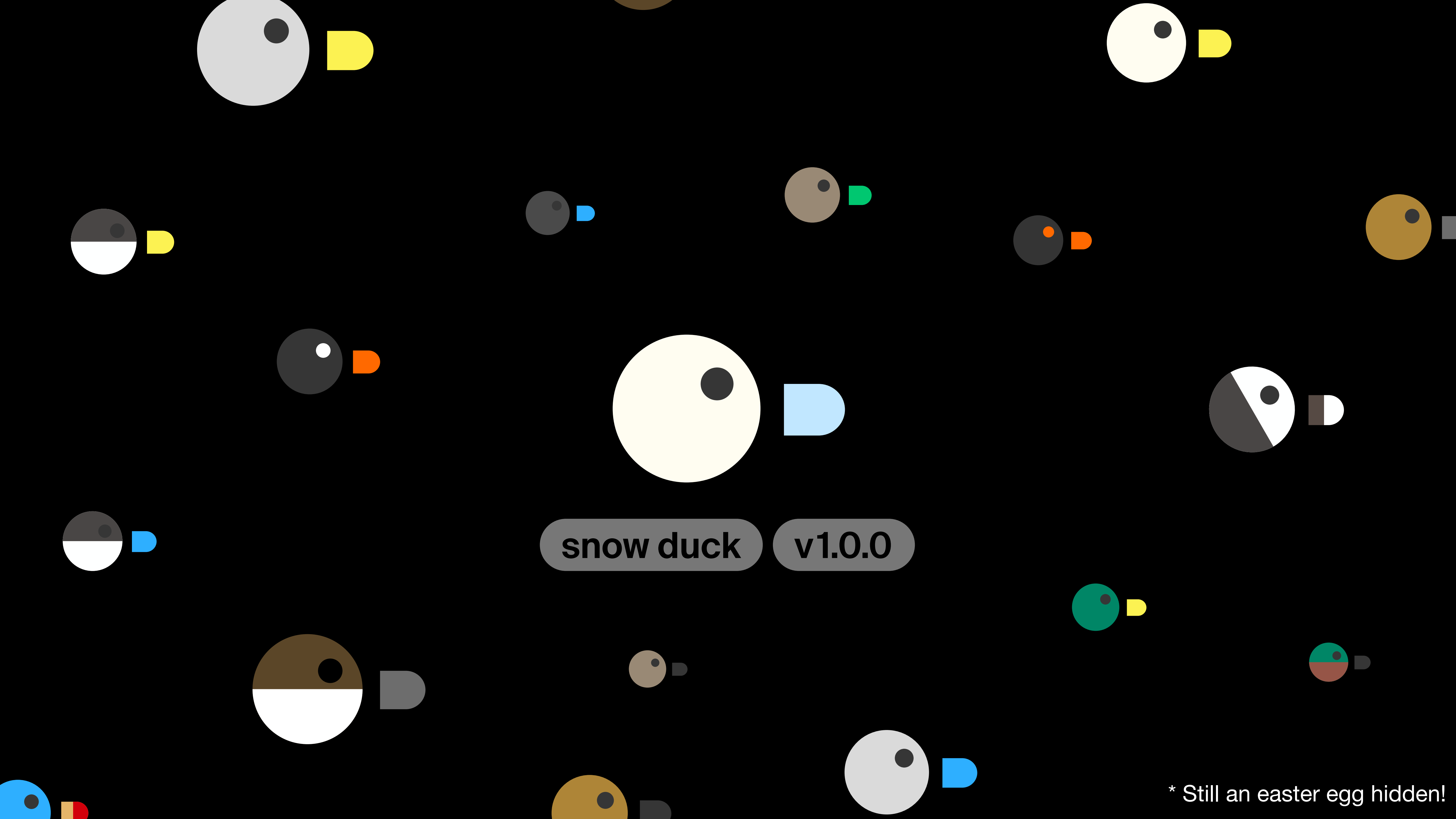


June 2023

January 2024

● Star History

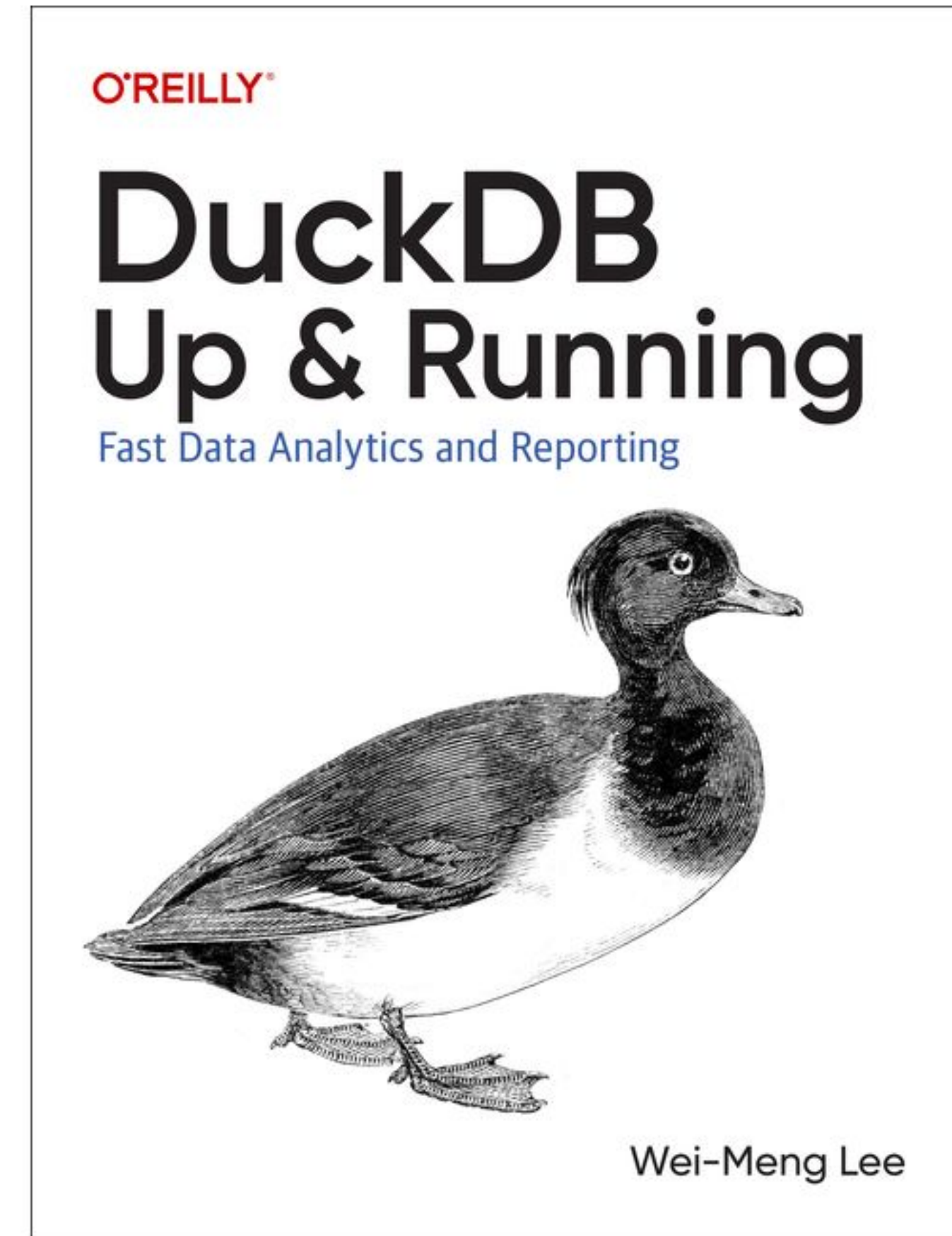
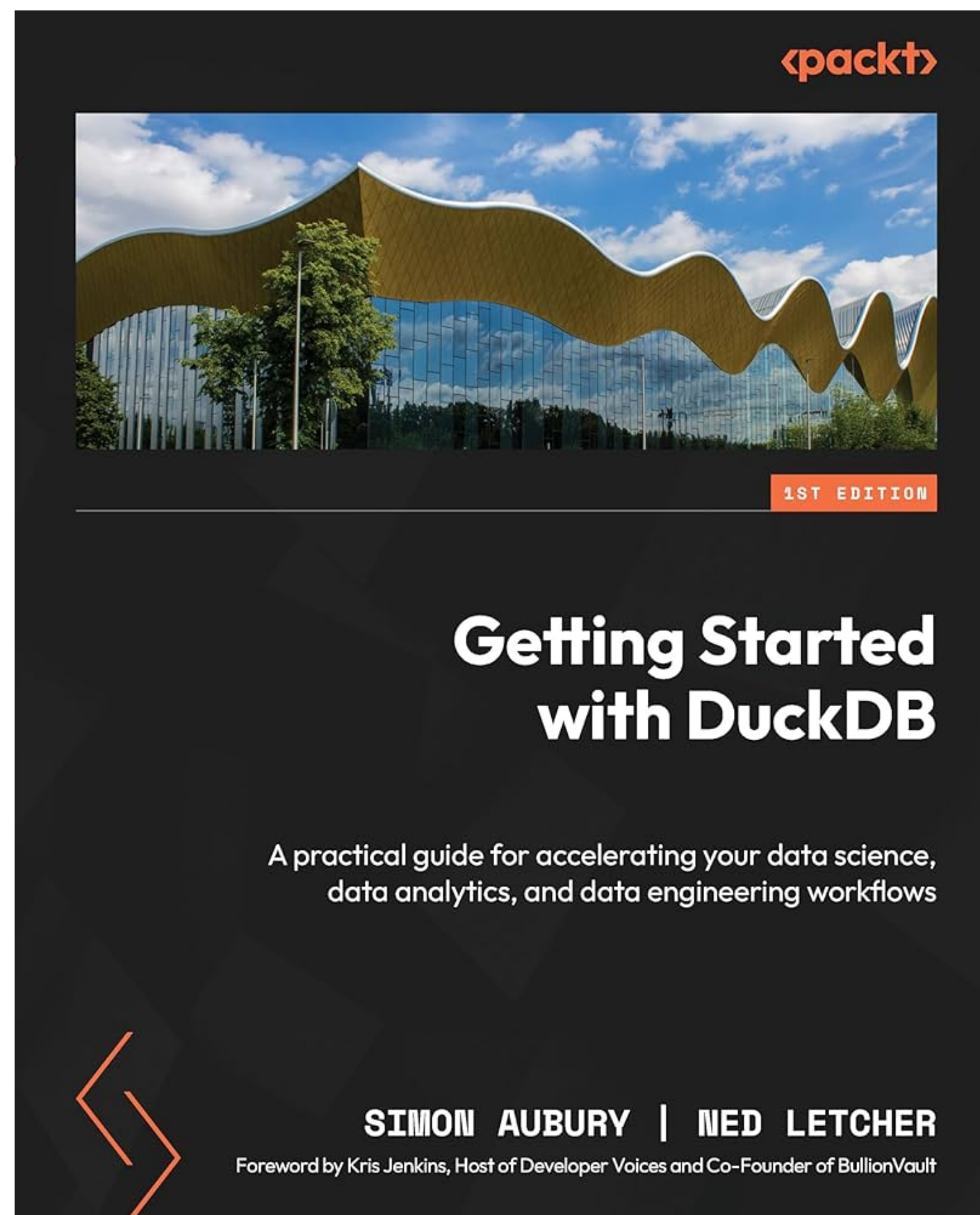
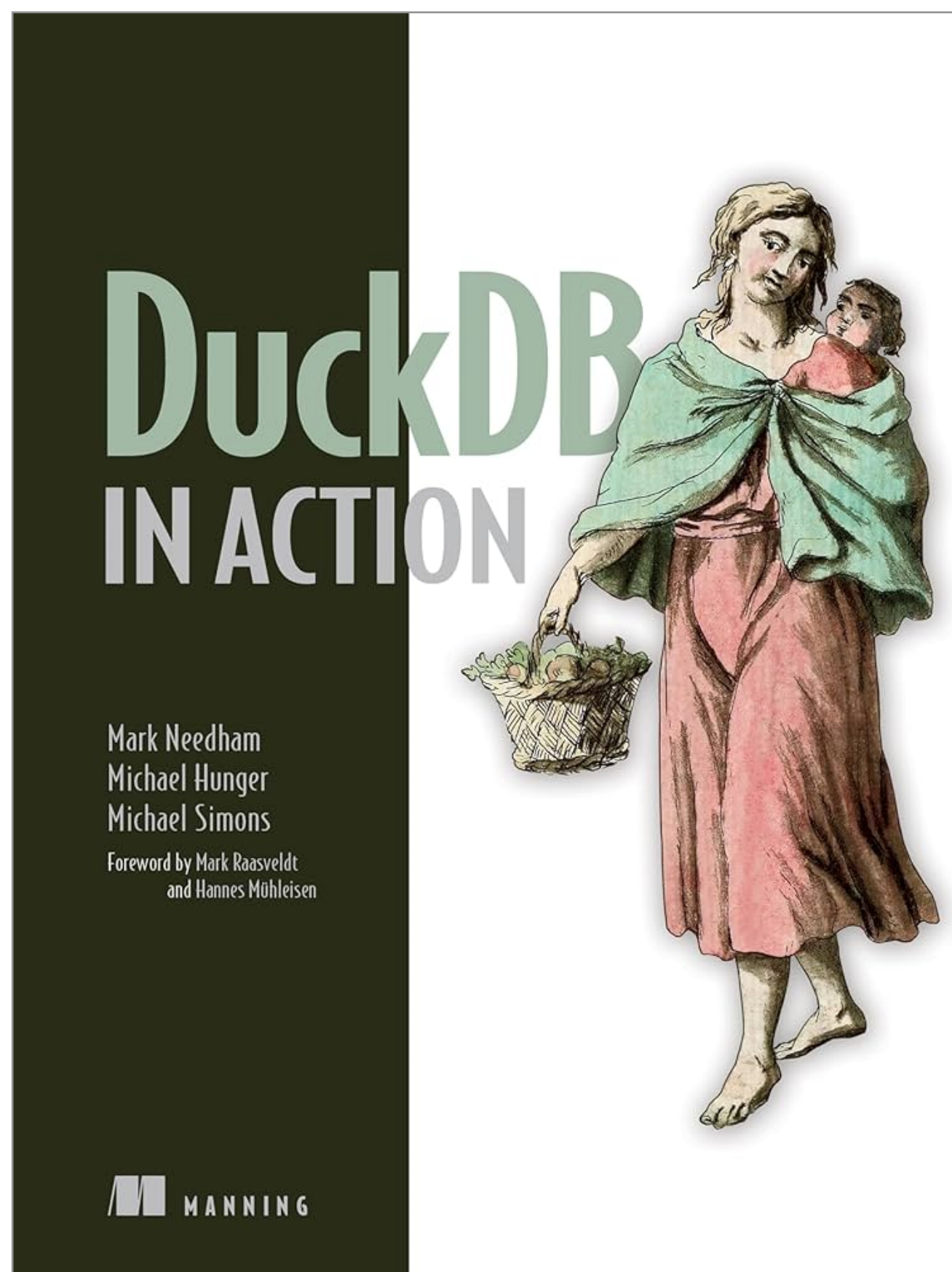




snow duck

v1.0.0

* Still an easter egg hidden!



⋮

DuckDB

Article

Talk

Read

Edit

View history

Tools

▼

From Wikipedia, the free encyclopedia

DuckDB is an [open-source column-oriented relational database management system](#) (RDBMS) originally developed by [Mark Raasveldt](#) and [Hannes Mühleisen](#) at the [Centrum Wiskunde & Informatica](#) (CWI) in the [Netherlands](#)^[1] and first released in 2019.^[2] The project has over 6 million downloads per month.^{[3][4][5]} It is designed to provide high performance on complex queries against large databases in embedded configuration,^[1] such as combining [tables](#) with hundreds of columns and billions of rows. Unlike other embedded databases (for example, [SQLite](#)) DuckDB is not focusing on transactional ([OLTP](#)) applications and instead is specialized for [online analytical processing](#) (OLAP) workloads ^[6]

DuckDB

Developer(s)	DuckDB Labs
Stable release	v1.0.0 / June 3, 2024
Repository	github.com/duckdb/duckdb ↗ ✎
Written in	C++
Operating system	Cross-platform
Type	Column-oriented DBMS RDBMS
License	MIT License

⋮

DuckDB

Article

Talk

Read

Edit

View history

Tools

▼

From Wikipedia, the free encyclopedia



This article **needs additional citations for verification**. Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed.

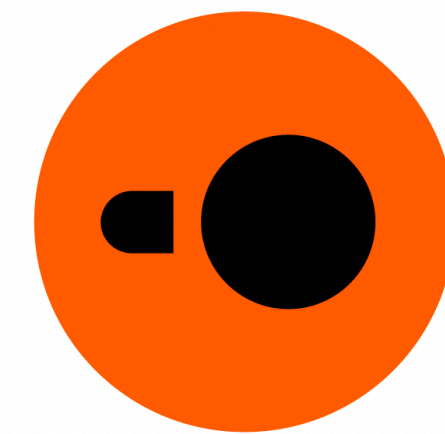
Find sources: "DuckDB" – news · newspapers · books · scholar · JSTOR (March 2024)
(Learn how and when to remove this message)

DuckDB is an open-source column-oriented relational database management system (RDBMS) originally developed by Mark Raasveldt and Hannes Mühleisen at the Centrum Wiskunde & Informatica (CWI) in the Netherlands^[1] and first released in 2019.^[2] The project has over 6 million downloads per month.^{[3][4][5]} It is

DuckDB

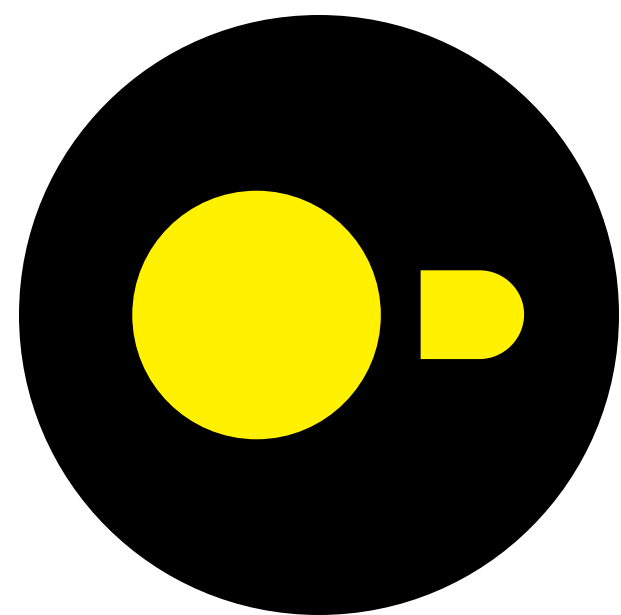
Developer(s)	DuckDB Labs
Stable release	v1.0.0 / June 3, 2024
Repository	github.com/duckdb /duckdb  

- **DuckDB Labs offers commercial support contracts**
 - Privileged access to DuckDB core team
 - Higher priority issues
 - Guaranteed response time
 - Private issue tracker
 - Private datasets
 - Extended support scope



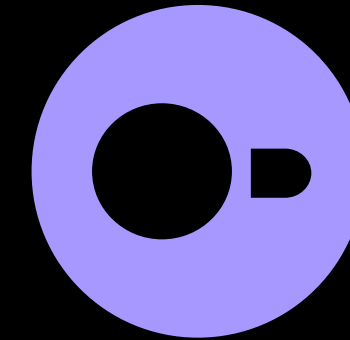
DuckDB Labs

<https://duckdblabs.com/support/>





- Where we are
 - Adoption
- **Where we're going**
 - Sneak Peek 1.1.0 “Eatoni”
- Q&A



DuckDB

[Overview](#)[Documentation](#)[Development](#)[FAQ](#)[Core Extensions](#)[List of Community Extensions](#)

Community Extension

Community Extensions

Welcome to the documentation for the DuckDB Community Extension Repository. This website contains all documentation specific to community extensions. For documentation on DuckDB itself, check out the [DuckDB docs](#).

What are Community Extensions

Community extensions are [DuckDB extensions](#) that are not maintained by the DuckDB team. This is different from the [Core extensions](#), which *are* maintained by the DuckDB team. Community extensions are distributed through the [Community Extension Repository](#), where anyone building an open-source DuckDB extension can make Pull Requests to submit an extension. Note that this makes DuckDB community extensions conceptually very similar to a package manager such as [Homebrew](#).

How to Use a Community Extension

To install and load a community extension, simply run:

```
INSTALL <extension_name> FROM community;  
LOAD <extension_name>;
```



community-extensions.duckdb.org

Community Extensions



Community Extension

h3

Hierarchical hexagonal indexing for geospatial data

[Extension repository on GitHub](#) ★ 146

[Extension descriptor ...](#) 

Installing and Loading

```
INSTALL h3 FROM community;  
LOAD h3;
```



Example

```
SELECT h3_latlng_to_cell(37.7887987, -122.3931578, 9);
```

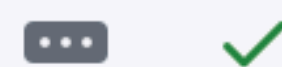




extension-template Public template

main

 **samansmink** Merge pull request [#77](#) from tshauck/improve-bootstrap

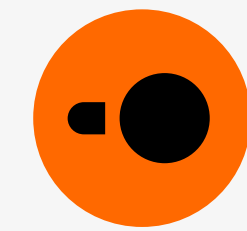


cb7641d · 3 weeks ago 207 Commits

 .github/workflows	add workaround for linux ci	3 weeks ago
 docs	cleanup and docs	6 months ago
 duckdb @ 1f98600	update to v1	2 months ago
 extension-ci-tools @ c0cc931	update to v1	2 months ago
 scripts	fix: make bootstrap check names earlier	last month
 src	Add version information	3 months ago

Code Blame 19 lines (17 loc)

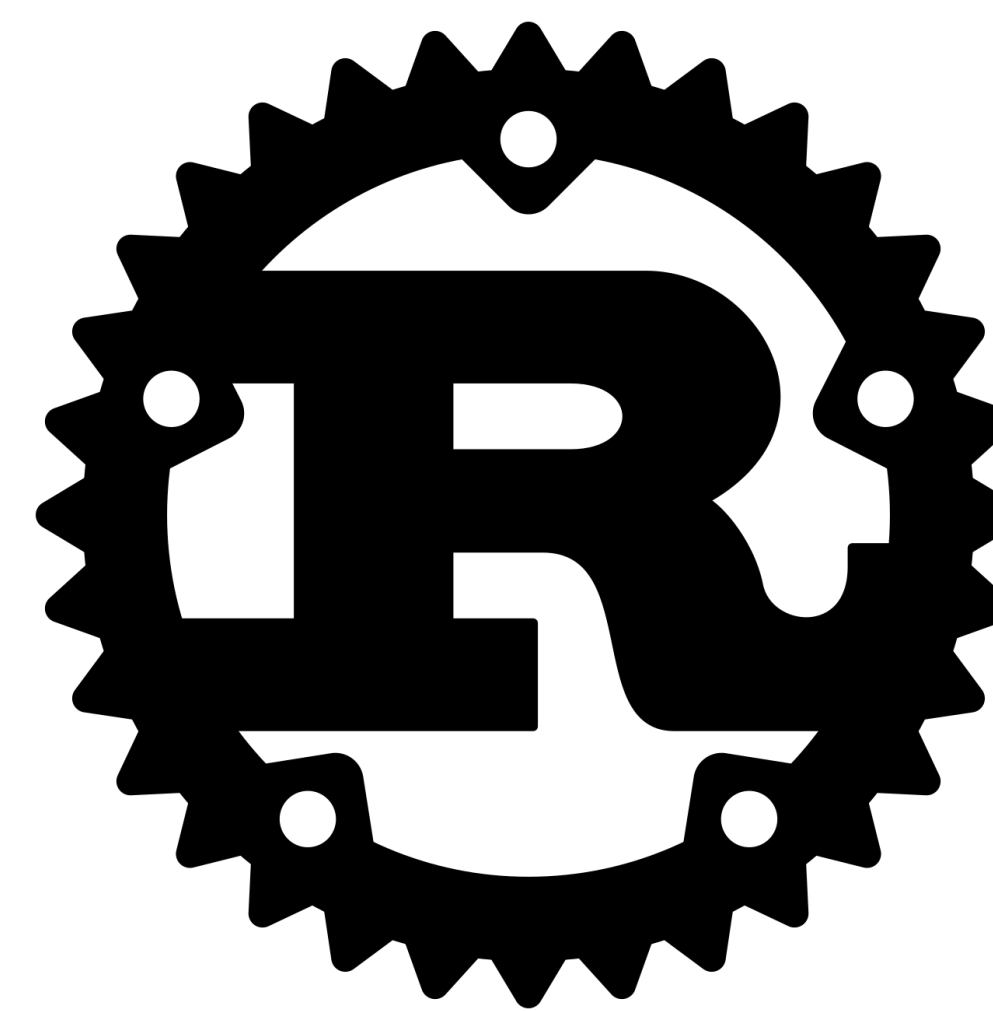
```
1  extension:
2    name: h3
3    description: Hierarchical hexagonal indexing for geospatial data
4    version: 1.0.0
5    language: C++
6    build: cmake
7    license: Apache-2.0
8    maintainers:
9      - isaacbrodsky
10
11  repo:
12    github: isaacbrodsky/h3-duckdb
13    ref: 8f2ebc1f09e9897f6ae1e1aaf9aca7a7b317d41a
14
15  docs:
16    hello_world: |
17      SELECT h3_latlng_to_cell(37.7887987, -122.3931578, 9);
18    extended_description: |
19      The H3 extension adds support for the [H3 hierarchical hexagonal grid system](https://h3geo.org/).
```

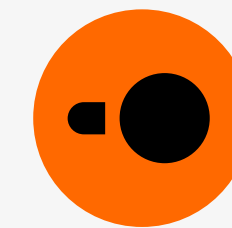


Current



Coming Soon





pg_duckdb: Official Postgres extension for DuckDB

pg_duckdb is a Postgres extension that embeds DuckDB's columnar-vectorized analytics engine and features into Postgres. We recommend using pg_duckdb to build high performance analytics and data-intensive applications.

pg_duckdb was developed in collaboration with our partners, [Hydra](#) and [MotherDuck](#).

DuckDB v1.1.0 “Eaton”



- **histogram** function allows you to quickly inspect the contents of a column
 - Bin boundaries determined automatically
 - Supported for all types

```
FROM histogram( '~/Data/ontime.parquet/*.parquet', uniquecarrier);
```

bin varchar	count uint64	bar varchar
AA	677215	
AS	139971	
B6	225718	
DL	696931	
EV	274565	
F9	75066	
NK	117714	
00	521956	
UA	435757	
WN	999114	
(other values)	112450	
11 rows		3 columns



- **query** and **query_table** functions

```
CREATE MACRO top_n_table(table_name) AS TABLE
SELECT MAX(COLUMNS(*), 3)
FROM query_table(table_name);

FROM top_n_values(nyctaxi);
```

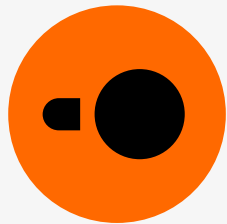
vendor_id varchar[]	pickup_at timestamp[]	...	tolls_amount float[]	improvement_surcha... float[]	total_amount float[]
[2, 2, 2]	[2015-12-31 23:59:...	...	[1901.4, 1450.09, ...	[137.63, 2.39, 1.43]	[3950611.5, 826040...
1 rows					19 columns (5 shown)

- SQL-level variables

```
SET VARIABLE files_to_read=(SELECT LIST(files) FROM file_list);
```

```
SELECT * FROM read_csv(getvariable('files_to_read')) LIMIT 5;
```

l_orderkey int64	l_partkey int64	l_suppkey int64	l_linenum int64	...	l_shipinstruct varchar	l_shipmode varchar	l_comment varchar
1	155190	7706	1	...	DELIVER IN PERSON	TRUCK	to beans x-ray car...
1	67310	7311	2	...	TAKE BACK RETURN	MAIL	according to the ...
1	63700	3701	3	...	TAKE BACK RETURN	REG AIR	ourts cajole above...
1	2132	4633	4	...	NONE	AIR	s cajole busily ab...
1	24027	1534	5	...	NONE	FOB	the regular, regu...
5 rows							
16 columns (7 shown)							



- Unpacked COLUMNS

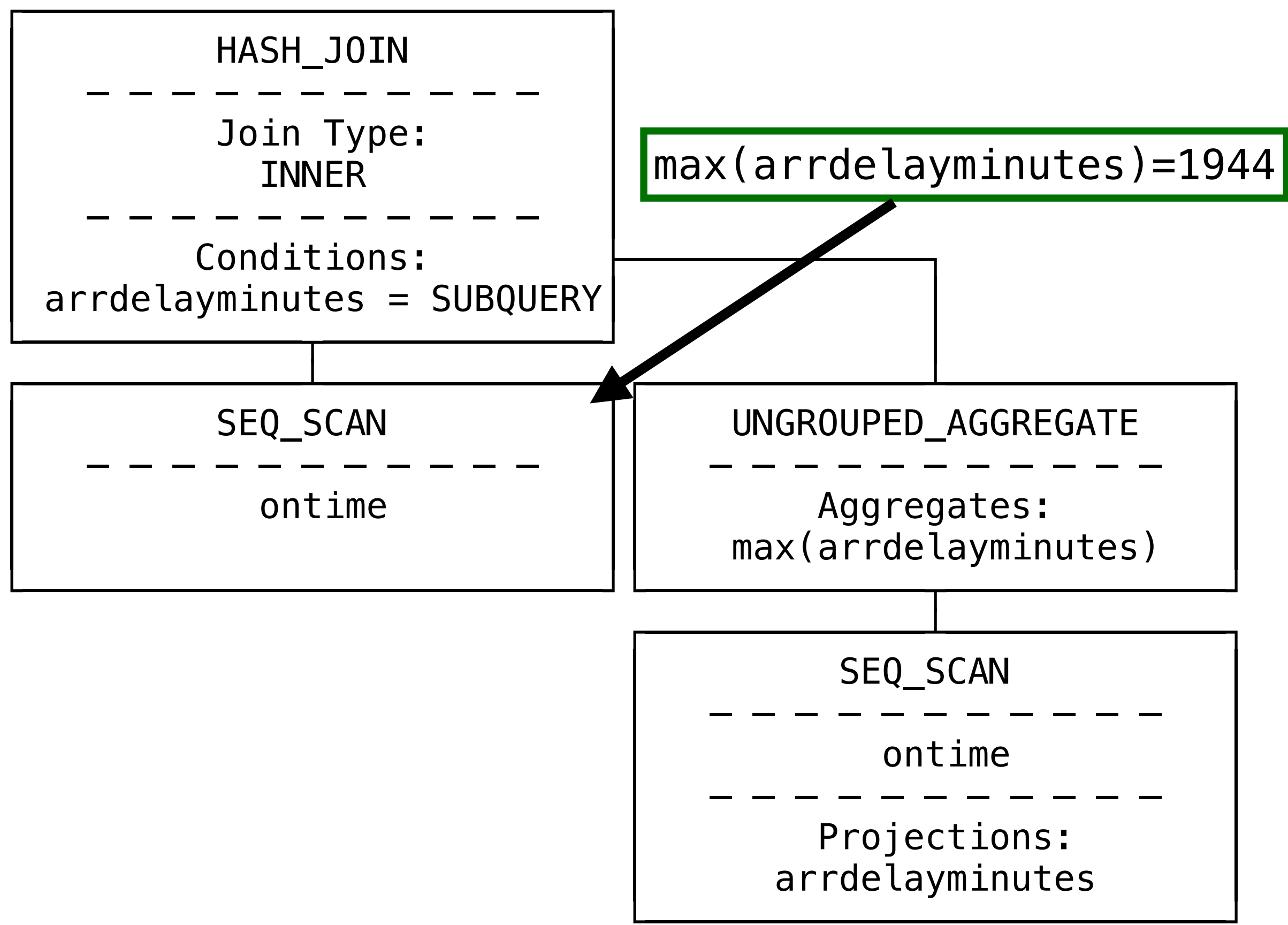
```
SELECT COLUMNS('arrdelay.*'), GREATEST(*COLUMNS('arrdelay.*')) AS greatest
FROM ontime
LIMIT 5;
```

arrdelay double	arrdelayminutes double	greatest double
-13.0	0.0	0.0
20.0	20.0	20.0
-14.0	0.0	0.0
-10.0	0.0	0.0
-43.0	0.0	0.0



- Dynamically pushdown filters from joins

```
SELECT *  
FROM ontime  
WHERE arrdelayminutes=(  
    SELECT max(arrdelayminutes)  
    FROM ontime  
);
```





- Many CSV reader improvements
- Adaptive sniffing - faster reading of many small CSV files
- Faster and less memory intensive union_by_name



file1.csv



file2.csv



file3.csv



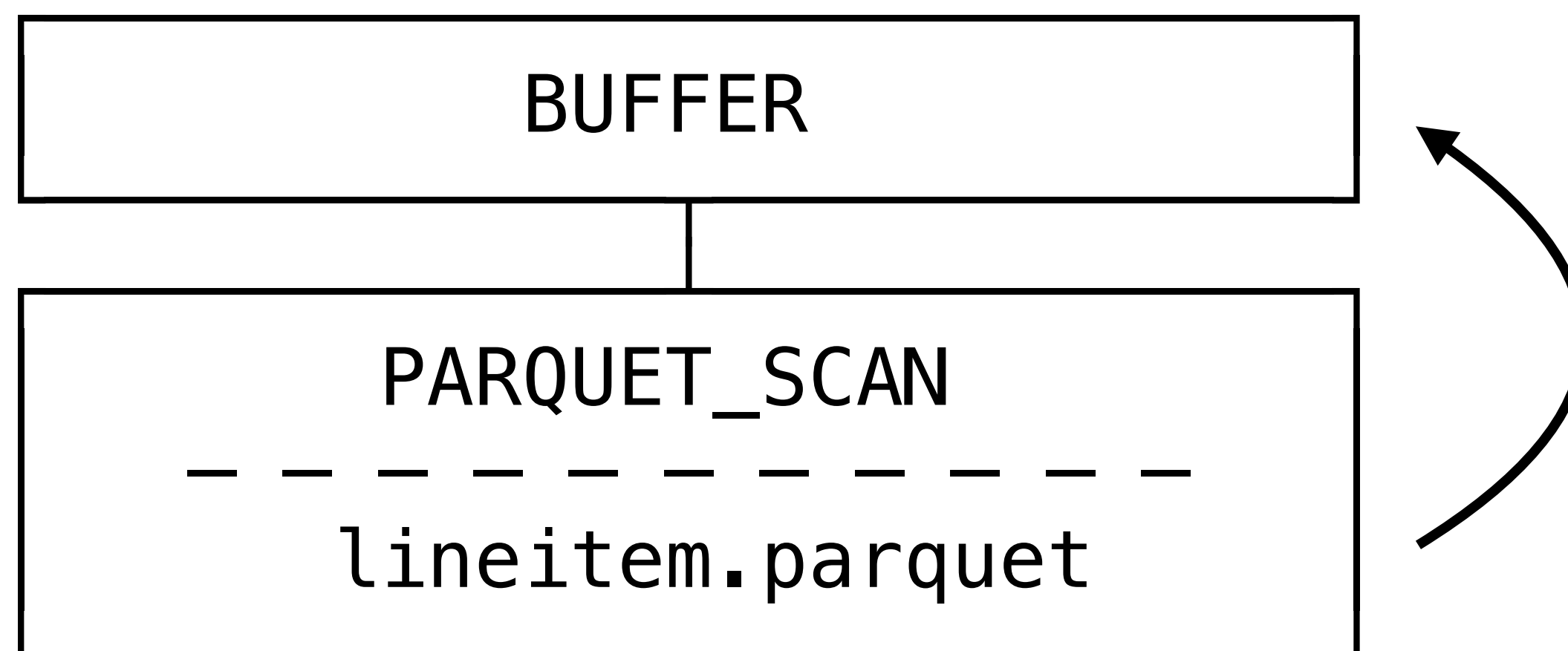
- Automatically re-use complex CTE results
- Avoids repeated expensive computation

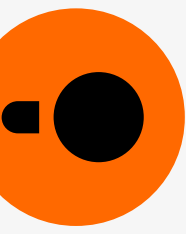
```
WITH revenue AS (  
    SELECT  
        l_suppkey AS supplier_no,  
        sum(l_extendedprice) AS total_revenue  
    FROM lineitem  
    GROUP BY supplier_no  
)  
SELECT *  
FROM supplier, revenue  
WHERE  
    s_suppkey = supplier_no  
    AND total_revenue = (  
        SELECT max(total_revenue)  
        FROM revenue)
```




- Parallel streaming queries
 - Data is read into buffers in parallel until buffer is full
 - When buffer is consumed parallel streaming is resumed
- High performance without requiring entire result to fit in memory

```
SELECT *  
FROM 's3:/duckdb-blobs/lineitem.parquet';
```



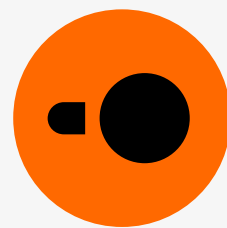


- GeoParquet read and write support

GeoParquet

Geospatial data in  Parquet

Documentation Dark Mode



 DuckDB

Q

Documentation ▾BlogGitHub★ 31.2k

Support

Q Search⌘+K

Documentation ▾

Installation

Data Import ▾

Client APIs ▾

SQL ▾

Introduction

Statements ▾

Data Types ▾

Extensions ▾

Docs / SQL

Light Modev0.10.0 (stable) ▾

SQL Introduction

Here we provide an overview of how to perform simple operations in SQL. This tutorial is only intended to give you an introduction and is in no way a complete tutorial on SQL. This tutorial is adapted from the [PostgreSQL tutorial](#).

In the examples that follow, we assume that you have installed the DuckDB Command Line Interface (CLI) shell. See the [installation page](#) for information on how to install the CLI. Launching the shell should give you the following prompt:

```
v0.9.2 3c695d7ba9
Enter ".help" for usage hints.
Connected to a transient in-memory database.
Use ".open FILENAME" to reopen on a persistent database.
D
```

IN THIS ARTICLE

SQL Introduction

Concepts

Creating a New Table

Populating a Table with Rows

Querying a Table

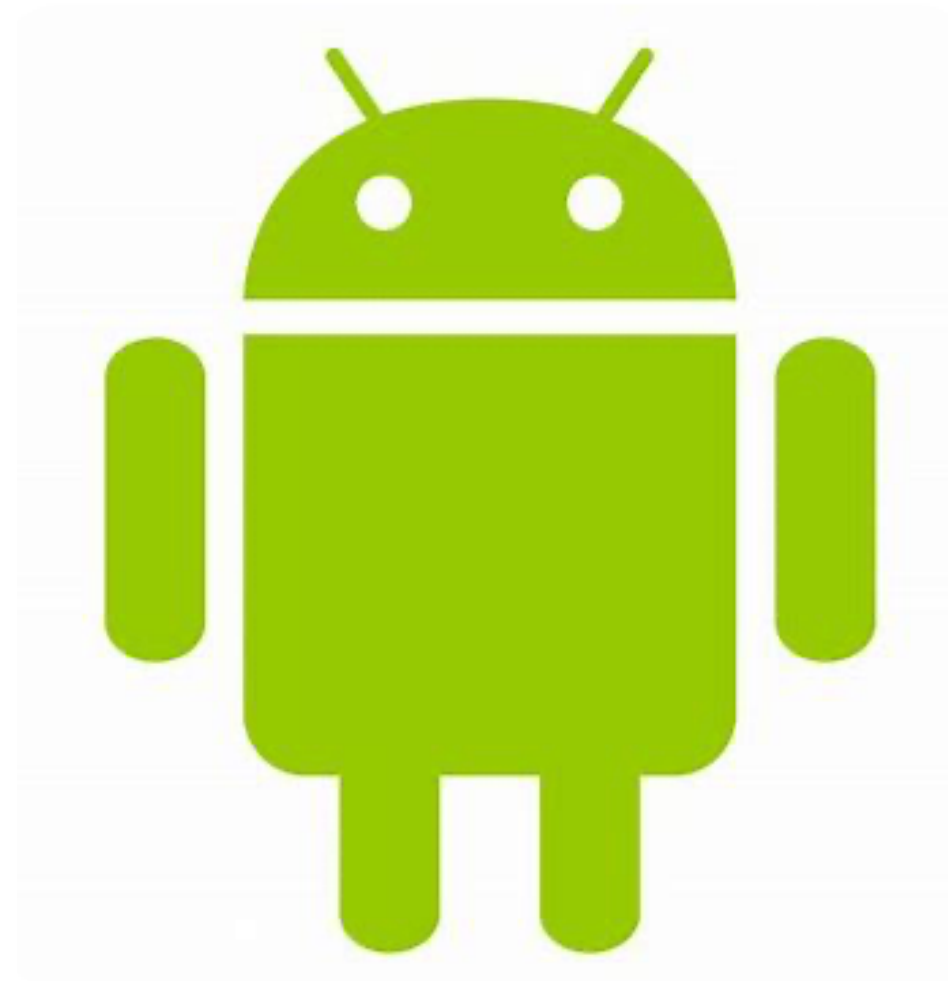
Joins between Tables

Aggregate Functions

Updates

Deletions

Platforms





- Extension Ecosystem
- Lakehouse Data Formats
- Optimiser Improvements
 - Partition/Sorting Awareness
 - Cardinality Estimation
- Extensible Parser



- Where we are
 - Adoption
- Where we're going
 - Sneak Peek 1.1.0
- **Q&A**

