

Making and Querying Databases in SQL with DuckDB

John Fink

McMaster University

November 26, 2025

Who am I?

Data in Abstract

- Data is a collection of facts and statistics

Data in Abstract

- Data is a collection of facts and statistics
- Can be raw, unorganized, or structured





What is notable/factual about a cat?

- Name

What is notable/factual about a cat?

- Name
- Sex

What is notable/factual about a cat?

- Name
- Sex
- Type or colouring

What is notable/factual about a cat?

- Name
- Sex
- Type or colouring
- Age

What is notable/factual about a cat?

- Name
- Sex
- Type or colouring
- Age
- Favourite food

What is notable/factual about a cat?

- Name
- Sex
- Type or colouring
- Age
- Favourite food
- Date acquired

Structured Data

- Data organized in a predefined format.

Structured Data

- Data organized in a predefined format.
- At its simplest, stored in tables with rows and columns.

Why Structured Data Matters

- Enables processing by computational means

Why Structured Data Matters

- Enables processing by computational means
- Can be manipulated in a regular fashion with SQL/other languages

Why Structured Data Matters

- Enables processing by computational means
- Can be manipulated in a regular fashion with SQL/other languages
- Ensures data portability and interoperability

CSV: Comma-Separated Values

- Simple text format for storing tabular data

CSV: Comma-Separated Values

- Simple text format for storing tabular data
- Each line represents a data record

CSV: Comma-Separated Values

- Simple text format for storing tabular data
- Each line represents a data record
- Fields are separated by commas

CSV: Comma-Separated Values

- Simple text format for storing tabular data
- Each line represents a data record
- Fields are separated by commas
- First line often contains column headers

CSV: Comma-Separated Values

- Simple text format for storing tabular data
- Each line represents a data record
- Fields are separated by commas
- First line often contains column headers
- Widely supported across platforms and tools

CSV: Comma-Separated Values

- Simple text format for storing tabular data
- Each line represents a data record
- Fields are separated by commas
- First line often contains column headers
- Widely supported across platforms and tools
- Human-readable and machine-readable

what is SQL?

- A *domain specific* language used to manage and query data

what is SQL?

- A *domain specific* language used to manage and query data
- Invented in 1973, last standardized in 2023

what is SQL?

- A *domain specific* language used to manage and query data
- Invented in 1973, last standardized in 2023
- Many different implementation exist, so a command valid for one database using SQL may not be valid for another.

What is DuckDB?

- an *open source, relational DB*

What is DuckDB?

- an *open source, relational DB*
- designed to provide *high performance* with complex queries on large datasets

What is DuckDB?

- an *open source, relational DB*
- designed to provide *high performance* with complex queries on large datasets
- yet itself is *small and portable*

What is DuckDB?

- an *open source, relational DB*
- designed to provide *high performance* with complex queries on large datasets
- yet itself is *small and portable*
- by default, works on simple, single files (no need to mess with accounts, access, etc)

What is DuckDB?

- an *open source, relational DB*
- designed to provide *high performance* with complex queries on large datasets
- yet itself is *small and portable*
- by default, works on simple, single files (no need to mess with accounts, access, etc)
- Runs on many, many operating systems, including an online version for testing and experimentation

What is DuckDB?

- an *open source, relational DB*
- designed to provide *high performance* with complex queries on large datasets
- yet itself is *small and portable*
- by default, works on simple, single files (no need to mess with accounts, access, etc)
- Runs on many, many operating systems, including an online version for testing and experimentation
- <https://duckdb.org/quickinstall>