PostgreSQL

powst · gres · kyoo · el

Neea Rusch 24 August 2022

Outline

1. Introduction $\sim$ 10 min
2. Technical demo $\sim$ 10-15 min
3. Research uses $\sim$ 5 min

Quick introduction

- Object-relational database management system (ORDBMS)
- Free and open-source software (FOSS)
- Developed by PostgreSQL Global Development Group
- Mature, widely-used and popular

Mature – many decades of history

1970s	Project **Ingres** developed at UC Berkeley
1980s	**Postgres**, the improved version of Ingres
1993	Development ended at UC Berkeley
1994	Open source community adopts **Postgres95**
1996	Final renaming to **PostgreSQL**

Widely-used and popular

Connection libraries available for many programming languages:
`C++`, `Java`, `Julia`, `Python`, `Node`, etc.

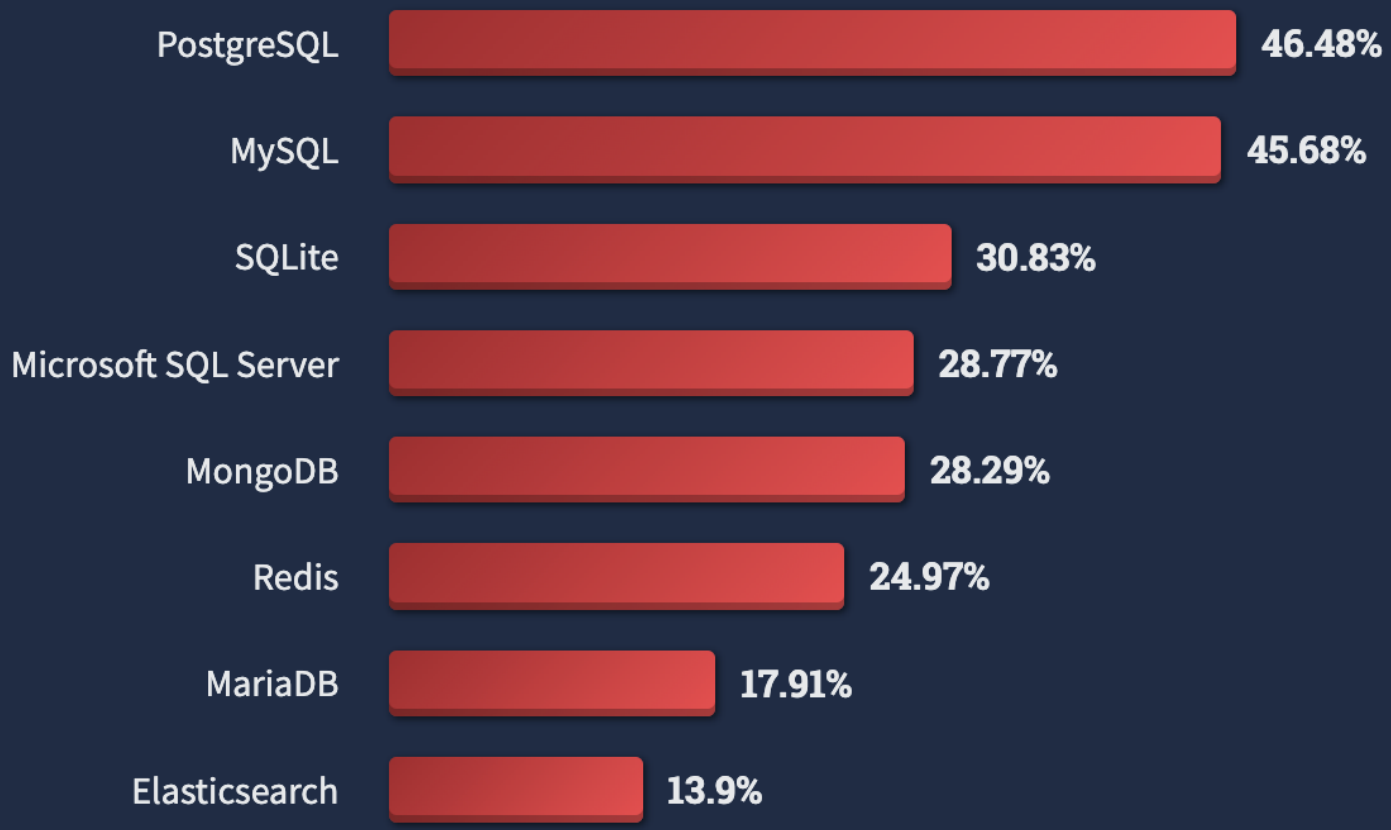
Over $158,000$ questions on Stack Overflow,
70% have accepted answers.

Some notable users:

Stack Overflow Developer Survey 2022

Q: Which database environments have you done extensive development work in over the past year, and which do you want to work in over the next year?

Professional developers – 48,788 responses



Features

All perks of relations databases

- schemas
- foreign keys
- flexible queries
- ACID transactions
- indices
- view tables
- stored procedures
- triggers, etc.

Features – data types

Natively supported types include:

arbitrary-precision numerics	arrays	boolean
character (text, varchar, char)	bit strings	binary
IPv4 and IPv6 addresses	composite	enum
date/time	JSON	money

Features – what makes it special

- Extensible by design
  extensions: GIS, statistics, text and image processing, ...
- Highly ANSI SQL-compliant
- Memory-efficient for performance-critical tasks
- Possible to process complex data types e.g., geographic data
- Object-oriented: tables inheritance
- Flexible full text search
- Built-in support for Unicode, localization, internationalization

Suitability assessment

Potential use cases:

- Structured data (ideally same attributes, dense)
- Can mix structured and unstructured data
- Scales up to multiple terabytes of data
- No fee, even for use in commercial software products

Not natively distributed, limited replication and offline/real-time sync support, excessive for small amounts of data.

Technical Demo

Usage basics

- Installation available for multiple platforms
  Linux, Windows, macOS, BSD, Solaris . . .
- Many ways to interact with database
  `psql` CLI is the primary front-end
- Default user is `postgres` and a password is not required
- Default port 5432

Demo notebook on Google Colab ↗

Research uses of PostgreSQL

```
SELECT *
FROM articles
WHERE body LIKE '%PostgreSQL%'
ORDER BY citations
LIMIT 3
```

Research uses

The #1 cited (56) paper[1]

- Researchers integrated decision tree data mining techniques into PostgreSQL
- Compared to an alternative tool (Weka), observed approx. 2-4x speedup in analysis times, for 100K - 1M records

[1] Viloria, Amelec, et al. *"Integration of data mining techniques to PostgreSQL database manager system."* Procedia Computer Science 155 (2019): 575-580.

Research uses

The #2 cited (45) paper[2]

- Premise: increased interest in content-based image retrieval
- Propose an image-handling extension: PostgreSQL-IE
- Able to create new feature extraction procedures, new feature vectors, and compose similarity queries
- Introduce a new image data type that can associate images

[2] Guliato, Denise, et al. *"PostgreSQL-IE: An image-handling extension for PostgreSQL."* Journal of digital imaging 22.2 (2009): 149-165.

Research uses

The #3 cited (33) paper[3]

"This report describes some characteristics of the development team of PostgreSQL that were uncovered by analyzing the history of its software artifacts as recorded by the project's CVS repository."

- User `momjian` (Bruce Momjian) is a key contributor
- Handful of core members, stable team

[3] German, Daniel M. *"A study of the contributors of PostgreSQL."* Proceedings of the 2006 international workshop on Mining software repositories. 2006.

Final remarks

PostgreSQL – expandable, flexible, open source ORDBMS

References

PostgreSQL documentation: `https://www.postgresql.org`

Perkins, Luc, et al. Seven Databases in Seven Weeks, 2nd Ed., 2018.

SO developer survey: `https://survey.stackoverflow.co/2022/`

Stackshare.io: `https://stackshare.io/postgresql`

Wikipedia `https://en.wikipedia.org/wiki/PostgreSQL`