

Landscape of Open Source Databases

Lorna Mitchell, Aiven



Keeping up with Databases

- We need more databases, because we have more data
- The right technology choice is important
- Open source is secureable and future-proof

Data Sources

Two main sources of data

- my own Opinion
- <https://db-engines.com/en/ranking>

Relational Databases

Traditional databases

- pre-defined tables with columns
- relations between tables, e.g. book has an author



MySQL

License: GPLv2

World's most-used open source database

- part of LAMP stack (Linux Apache MySQL PHP/Python/Perl)
- proprietary Enterprise Server version also available



MariaDB

License: GPLv2

- drop-in replacement for MySQL
- support for additional storage engines
- proprietary Enterprise Server version also available



PostgreSQL

License: PostgreSQL license (MIT-ish)

- powerful and performant relational database
- many contributors, healthy community
- lots of extensions



PostGIS

License: GPLv2

Spatial database, as an extension to PostgreSQL.

- support for geographical object data types
- functions for working with area, distance, etc
- specialist indexes to support spatial queries



TimescaleDB

License: Apache2, some features TSL

Extension for PostgreSQL

- table types for timeseries data
- additional SQL functions



Time Series Data

Time series data:

- a timestamp
- a measurement



InfluxDB

License: MIT

- time series database
- IoT, metrics, energy
- clustered version has proprietary license



Re-use wire protocols

Build a new database, use an existing wire protocol to get clients and integrations

Examples:

- CrateDB uses PostgreSQL protocol
- VictoriaMetrics and M3DB use Influx and Prometheus protocols

SQLite

License: public domain

- file based, no server
- embeddable
- ideal edge model database



Redis

License: BSD

Speedy in-memory key value store

- used for caching, queueing
- supports many data types (lists, sets, hashes, etc)
- 3rd most popular open source database



Key/Value Stores

Other key value stores worth a mention:

- Memcached
- etcd
- ArangoDB



Apache Cassandra

License: Apache2

- distributed database for commodity hardware
- designed for very large volumes of data
- use denormalised data storage (no joins)



Distributed Databases

Horizontally scalable for writes, spread across multiple nodes

- data organised into shards or partitions
- usually also replicated for redundancy
- complexity handled by database



OpenSearch

License: Apache2

Open source fork of Elasticsearch

- powerful search and aggregation features
- flexible data structure, but defined indexes
- Opensearch Dashboards is the fork of Kibana

Open Source Databases

Best technology around, whatever your data needs

Resources

- <https://aiven.io> - DBaaS
- <https://uptime.aiven.io> - Open source data event
- <https://lornajane.net> - my website/blog
- 7 Databases in 7 Weeks (2nd edition)