# The Open Source Licenses of Popular DBMSs

Christian Weckner

January 17, 2023

**Abstract**

Database Management Systems have been a critical part of digital infrastructure for decades. They are a necessary tool to efficiently work with large amounts of data. Through the years, as the way data is used has changed, so has the DBMSs. Today, there are hundreds to choose from, each with pros and cons, support for different data models. As they are just pieces of software, some are use commercial licenses, while others use Open Source licenses. On DB-Engines.com, there is a ranking of the most popular DBMSs, based on certain criteria. 18 of these were chosen, and had their licenses analyzed. Out of the 18, 12 were licensed as Open Source. Breaking the Open Source DBMSs down further, 8 used permissive licenses, 2 used copyleft licenses, and 2 used licenses which are actually disputed. The Open Source Initiative have criticized the two new licenses, called the Server Side Public License and the Elastic License. These were created as a response to the growth of the Software as a Service-industry, and claims to combat the theft Open Source software. The response to these licenses have been fairly negative, despite their claims of openness. How they will affect the DBMS landscape, and Open Source software in general, remains to be seen.

# Acknowledgements

# 1 Introduction

A Database Management System (DBMS) is a software used for managing and storing data. DBMSs have been around for decades, and are an integral part of the digital infrastructure in society. Since DBMSs are pieces of software, they have inevitably been in contact with the Open Source software movement, and thus have been licensed under traditional Open Source licenses. As the way software is being used and provided changes, the licenses must adapt.

This study will look at some of the most popular DBMSs, and what Open Source licenses they use. It will also look at which DBMSs have changed licenses, why they did it, and what the response was.

# 2 Background

## 2.1 Databases and Database Management Systems

Though both are commonly referred to as databases, a database and a Database Management System are not the same thing. A database is a collection of structured data. Since data comes in many different shapes, databases can have different structures. Most common is the traditional rows and columns approach, like a spreadsheet. A database entry is represented as a row, with the columns being the different fields of the entry. With this structure, all entries have the same shape. However, this kind of rigid structure does not fit all use cases. In some cases, the need for more 'columns' than defined in the schema might be necessary, or perhaps 'columns' with different names. This can be done in a spreadsheet-style database, but not without complex queries. Instead, databases with different structures have been developed to better fit these use cases.

A Database Management System is the software used to manage a database. Modern DBMSs are needed to

- handle persistent data,

- give efficient access to large amounts of data,

- guarantee integrity constraints of data, and

- handle transactions and concurrent access to data.

Just like there are different kinds of databases, i.e. ways to structure data, there are different kinds of DBMSs. Some, like Neo4j, specialize in a specific model (graph databases for Neo4j). Others, like MySQL have a primary model (relational), but also supports multiple different models.

## 2.2 DB-Engines

DB-Engines is a website which collect and present information on database management systems. DB-Engines have a monthly ranking, containing 399 different DBMSs, ranked from most to least popular. This ranking is based on factors such as mentions in job listings and forum pages. [9]

## 2.3 Open Source Licenses

As a Database Management System is a piece of software, it can be licensed using an Open Source license. These licenses gives the author legal protection, and tells the user how the code may be used, modified and redistributed. Open Source licenses are divided into three categories:

**Public Domain**: When choosing to release code under Public Domain, or any equivalent license, the author surrenders any copyright. The users are free to do anything and everything with the code, at their own risk. An example of a Public Domain license is: CC-0. [1]

**Permissive**: Permissive licenses, in general, provides the same user rights as a Public Domain license, but with the requirement of properly attributing the author of the original work. Some examples of permissive licenses are: Apache 2.0, BSD3c, MIT, CC-BY. [22]

**Copyleft**: The copyleft licenses are the most restrictive licenses. The key difference from more permissive licenses is that copyleft licenses prohibit propreitization of the code in question. If copyleft code is modified and distributed, then the modified code must also be licensed and openly released under a compatible license. Some examples of copyleft licenses are GPLv3, AGPL.[23]

# 3 Data

## 3.1 The Database Management Systems

The DBMSs were chosen using the popularity ranking [9] available on DB-engines. This ranking was used because DB-engines clearly states which factors are taken into account when calculating the popularity score. Since there are 399 DBMSs listed in the ranking, some constraints were introduced:

- Only models whose top ranked DBMS were in the top 20 most popular were considered. This resulted in 6 different models.

- Only the top 3 of each model were considered. This resulted in a total of 18 DBMSs.

## 3.2 Gathering the data

Ironically, as DB-engines does not provide an API, the data was gathered manually in October 2022. Each DBMS-model has its own ranking, and from this, the top 5 were collected. If the top ranking DBMS was not in the top 20 overall, then the entire model was discarded.

In the remaining models, there were some overlap, I.E. a DBMS appeared in multiple rankings. This is due to the concept of multi-models; a DBMS can be used in multiple ways. In response to this, if a DBMS was present in multiple models, then it would be chosen to represent the model in which it ranked the highest.

E.g. if **DBMS A** was ranked # 2 in *model B*, and #3 in *model C*, then it would be included in *model B*'s top 3. For *model C*, the next DBMS to consider is the one ranked #4.

As DB-engines doesn't always list all licenses previously used, additional research was performed to find if and when licenses had been changed.

The final dataset used is available on GitHub. [21]

# 4 Results

## 4.1 The DBMSs

The final models and DBMSs were:

- **Relational**: Oracle, MySQL[16], Microsoft SQL

- **Document Store**: MongoDB[15], Amazon DynamoDB, Databricks[7]

- **Key-Value Store**: Redis[18], Memcached[14], Hazelcast[10]

- **Search Engine**: Elasticsearch[8], Splunk, Solr[19]

- **Wide-Column Store**: Cassandra[6], HBase[11], Datastax

- **Graph**: Neo4j[17], Microsoft Azure Cosmos DB, ArangoDB[2]
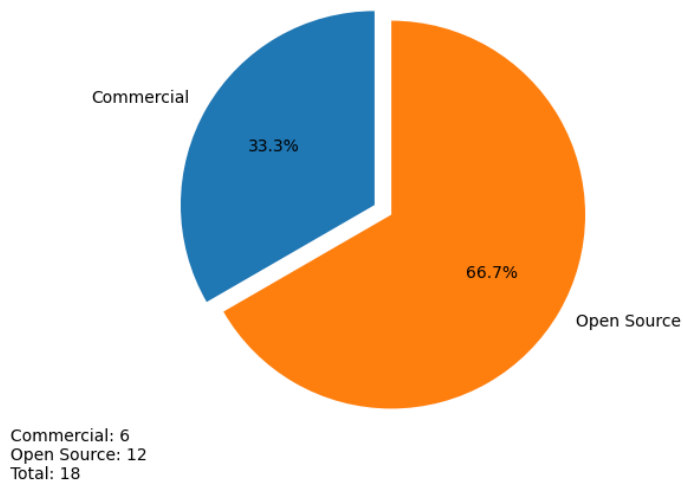
## 4.2 Open Source or Commercial



Figure 1: Number of Open Source databases in relation to Commercial

Of the 18 DBMSs, 12 had an Open Source License.
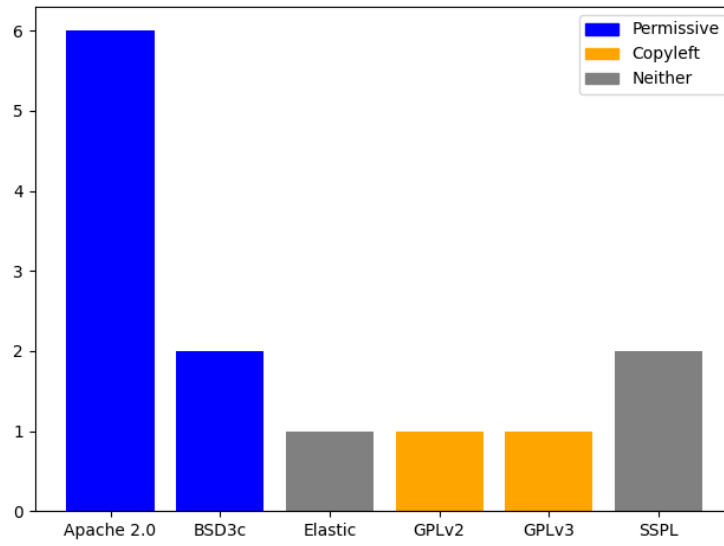
## 4.3 Permissive or Copyleft



Figure 2: The Open Source licenses represented, and their count.

6 different licenses were represented. The total count is 13, not 12, as ElasticSearch is dual-licensed under both Elastic License v2 and SSPL. 8 DBMSs used a *permissive* license, 2 used a *copyleft* license. The remaining two DBMSs (MongoDB and ElasticSearch) used neither.

## 4.4 Relicensing

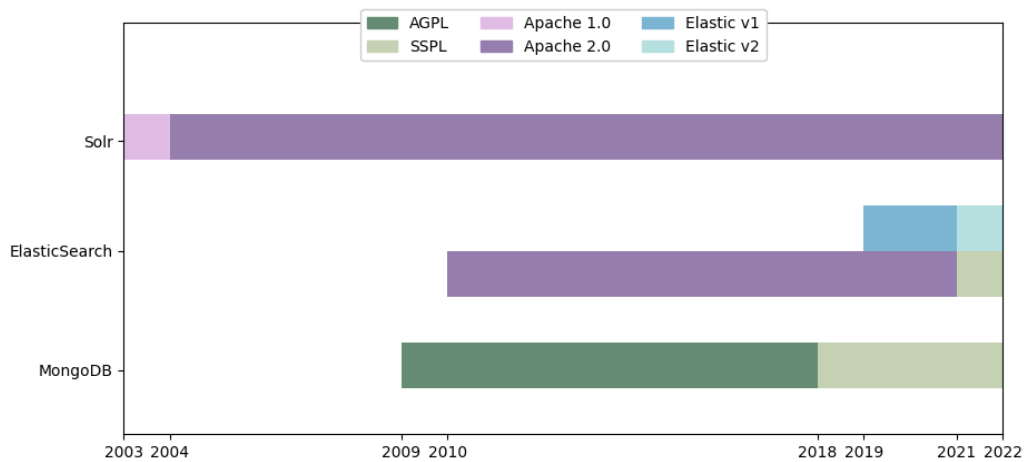Out of the 13 Open Source DBMSs, 3 have at some point relicensed.



Figure 3: A timeline of the DBMSs, the new and old licenses, and the year of relicensing.

### 4.4.1 Solr

Solr, being developed by Apache, was released in 2003 under Apache 1.0, and naturally relicensed to Apache 2.0 upon its release in 2004.

### 4.4.2  MongoDB

Upon release in 2009, MongoDB was licensed under the Affero GNU Public License. AGPL is a copyleft license based on GPLv3, with an added section about the use of modified licensed software through a computer network. In 2018, MongoDB was relicensed under the Server Side Public License, a license created by MongoDB, Inc.. SSPL is similar to AGPL, as both are derivatives of GPLv3, with the difference being Section 13.

In AGPL, Section 13 states that if a AGPL-licensed program is modified, and can be interacted with remotely through a computer network, then the modified code must be made available to the user easily and free of charge. The full Section 13 of AGPL can be read in appendix A. MongoDB, Inc. claims that the wording "interacted with" is too vague, and has been circumvented by large service providers. To combat this, SSPL was created with "clearer" conditions. These new conditions are, in short, that if a SSPL program, or a modified version of it, is used anywhere in the service, then all source code of the service must be released. Section 13 of SSPL can be read in appendix B.

As an example, if a company called BongoDB, Inc. wrote a proprietary API which talks to an AGPL-licensed version of MongoDB, they could sell it as a service, without triggering Section 13 of AGPL. The reason for this is that the MongoDB code is unmodified, and already available through MongoDB, Inc. The API is a separate piece of code, which just happens to give access to MongoDB functionality. However, if another company, called PongoDB, Inc. wrote a proprietary API which instead talks to an SSPL-licensed version of MongoDB, Section 13 would trigger, and force PongoDB, Inc. to release their API-code as Open Source.

### 4.4.3  ElasticSearch

On initial release in 2010, ElasticSearch was licensed under Apache 2.0. However, in 2019, Elastic introduced their own license called Elastic License v1. In 2021, both licenses were replaced with SSPL and Elastic License v2, respectively.

Elastic License v2 is described by Elastic as *non-copyleft*, and gives the user the right to "use, copy, distribute, make available, and prepare derivative works of the software", provided that certain limitations are followed. These can be read in Appendix C. Specifically, the Elastic License v2 forbids the use of Elastic License v2-code being included in SaaS-solutions. Elastic License v1 was created after an incident with Amazon Web Services. AWS violated Elastics trademark by offering a service called Amazon Elasticsearch Service, saying that it was in collaboration with Elastic. Elastic put up a blog post calling this out, stating this as the reason for the new license. [4]

SSPL was chosen as a dual license simply to mitigate the effects that the change would have on their users. [5]

# 5    Discussion

## 5.1    Summary of results

Out of the 399 DBMSs listed on DB-Engines, 18 were selected for this study. From this dataset, 12 were licensed under an Open Source license. 8 of these were licensed with a permissive license, 2 with a copyleft license, and 0 with a public domain license. The final 2 DBMSs were licensed using licenses not generally considered Open Source, even though they are listed as such on DB-Engines.

## 5.2    The Case of SSPL and Elastic

While both SSPL and Elastic License stems from Open Source projects, neither have been accepted as Open Source licenses by organizations such as the Free Software Foundation or the Open Source Initiative. MongoDB, Inc. tried to get SSPL approved by the OSI in 2019, but later withdrew the application.[12] The reason for this was a dispute regarding Section 13 of SSPL, which would interfere with Section 6 and 9 of the Open Source Definition [3], a list of criteria which Open Source software must fulfill. Specifically, SSPL would force "Software as a Service"-vendors to release their entire service stack, which would possibly interfere with the licenses of the software used in the stack (Section 9 of OSD). This in turn would make it more difficult to provide the SaaS (Section 6 of OSD).[20]

When Elastic relicensed ElasticSearch under SSPL and Elastic License v2, the OSI responded with a blog post [13], sharing their concerns about the licenses. While Elastic License v2 was never submitted to the OSI, they still mention that the license does not comply with Section 6 of the OSD, much like SSPL.

The OSI are not the only one criticizing these new licenses. The main points which cause friction are the claims of openness despite restricting SaaS, and that the power is now in the hands of the developers of the SSPL-software. SSPL would give them more control of pricing, forcing the vendors to either have to use enterprise licenses, which would make the service more expensive, or to release the entire service stack, which might not even be possible. [25] [26] [24]

For all the points brought up against SSPL and Elastic License, there is a possible positive outcome which is skipped over. Just like in a forest, when an old, tall tree falls, it allows the plants at the forest floor to get some sunlight and grow stronger, improving the biodiversity of the area. The same could happen if SSPL licensed products end up being unattractive for SaaS-vendors. There is no shortage of DBMSs that could replace MongoDB or ElasticSearch.

The creation of SSPL and Elastic License also raises the question of whether AGPL is enough? The license is 15 years old, and from a time when SaaS was nowhere near as popular as it is today.[**SaaS-trends**] If the claimed circumvention of AGPL continues, then maybe the license needs a revision? There is also the possibility that others fork GPLv3, in an attempt to create a "better" SSPL. Currently, the Free Software Foundation has not released anything regarding an update AGPL, but if the use of SSPL keeps growing, then the FSF might have to respond.

## 5.3    Future Work

Performing this study raised some questions which are not presented in the final paper, as they would work better as their own projects. Some of these are:

- How has SSPL affected the landscape?
  – Which DBMSs have relicensed under SSPL?
  — Which licenses did they go from?
  – Has the relicensing affected their ranking?
  — ... community contributions?
  — ... economic results?
  – Has it affected the DBMSs in the same category who did not relicense?
  – Has the relicensing caused SaaS-vendors to drop the DBMS from their service stack?
  — Which DBMSs were chosen as replacement?


- A case study of the top 5 Open Source DBMS with regard to...
  – ... community size and contributions.

- ... business plan.
  - ... and other Open Source practices.


- An expanded version of this study, with all 399 DBMSs, looking at...
  - ... commercial vs Open Source.
  - ... permissive vs copyleft vs public domain vs neither.
  - ... ranking trends for the different licenses, or license types.


## 5.4    Conclusions

Not only was Open Source licenses more common in the top than commercial licenses, the number of permissive licenses outnumbered the copyleft licenses 3-to-1. There were no public domain licenses present at all in the dataset. This was not surprising, as the developers most likely wants some credit for their work.

The study found two "new" licenses, SSPL and Elastic License, which are not accepted as Open Source by large governing bodies such as the Free Software Foundation and the Open Source Initiative. It's still unclear if these licenses will have an impact on the landscape, and what the effects will be, but they are a sign that as the way we use software change, the licenses which tells us how we may use the software perhaps has to change as well.

# References

[1]  *About the Licenses.* unavailable. URL: https://creativecommons.org/licenses (visited on 01/16/2023).

[2]  *ArangoDB GitHub.* URL: https://github.com/arangodb/arangodb (visited on 01/16/2023).

[3]  Lukas Atkinson. *[License-review] February 2019 Summary.* 2019. URL: http://lists.opensource.org/pipermail/license-review_lists.opensource.org/2019-March/003988.html (visited on 01/16/2023).

[4]  Shay Banon. *Amazon: NOT OK - why we had to change Elastic licensing.* 2021. URL: https://www.elastic.co/blog/why-license-change-aws (visited on 01/16/2023).

[5]  Shay Banon. *Introducing Elastic License v2, simplified and more permissive; SSPL remains an option.* 2021. URL: https://www.elastic.co/blog/elastic-license-v2 (visited on 01/16/2023).

[6]  *Cassandra GitHub.* URL: https://github.com/apache/cassandra (visited on 01/16/2023).

[7]  *Databricks GitHub.* URL: https://github.com/databricks (visited on 01/16/2023).

[8]  *ElasticSearch.* URL: https://github.com/elastic/elasticsearch (visited on 01/16/2023).

[9]  DB-Engines. *DB-Engines Ranking.* 2022. URL: https://web.archive.org/web/20221019052055/https://db-engines.com/en/ranking (visited on 01/16/2023).

[10]  *Hazelcast GitHub.* URL: https://github.com/hazelcast/hazelcast (visited on 01/16/2023).

[11]  *HBase GitHub.* URL: https://github.com/apache/hbase (visited on 01/16/2023).

[12]  Eliot Horowitz. *[License-review] Approval: Server Side Public License, Version 2 (SSPL v2).* 2019. URL: http://lists.opensource.org/pipermail/license-review_lists.opensource.org/2019-March/003989.html (visited on 01/16/2023).

[13]  Open Source Initiative. *The SSPL is Not an Open Source License.* 2021. URL: https://opensource.org/sspl-not-open-source (visited on 01/16/2023).

[14]  *Memcached GitHub.* URL: https://github.com/memcached/memcached (visited on 01/16/2023).

[15]  *MongoDB GitHub.* URL: https://github.com/mongodb/mongo (visited on 01/16/2023).

[16]  *MySQL GitHub.* URL: https://github.com/mysql/mysql-server (visited on 01/16/2023).

[17]  *Neo4j GitHub.* URL: https://github.com/neo4j/neo4j (visited on 01/16/2023).

[18]  *Redis GitHub.* URL: https://github.com/redis/redis (visited on 01/16/2023).

[19]  *Solr GitHub.* URL: https://github.com/apache/solr (visited on 01/16/2023).

[20]  *The Open Source Definition.* 2007. URL: https://opensource.org/osd (visited on 01/16/2023).

[21]  Christian Weckner. *Dataset of DBMSs.* 2023. URL: https://github.com/cweckner/OpSci_DB_Dataset (visited on 01/16/2023).

[22]  *What is a "permissive" Open Source license?* unavailable. URL: https://opensource.org/faq#copyleft (visited on 01/16/2023).

[23]  *What is Copyleft?* unavailable. URL: https://opensource.org/faq#copyleft (visited on 01/16/2023).

[24]  Peter Zaitsev. *The Case Against the Server Side Public License (SSPL).* 2022. URL: https://thenewstack.io/the-case-against-the-server-side-public-license-sspl/ (visited on 01/16/2023).

[25]  Peter Zaitsev. *Why is MongoDB's SSPL Bad For You?* 2020. URL: https://www.percona.com/blog/2020/06/16/why-is-mongodbs-sspl-bad-for-you/ (visited on 01/16/2023).

[26]  Peter Zaitsev. *Why SSPL is Bad For You, Part 2.* 2021. URL: https://www.percona.com/blog/2021/02/02/why-sspl-is-bad-for-you-part-2/ (visited on 01/16/2023).

# A    GNU Affero General Public License Section 13

13. Remote Network Interaction; Use with the GNU General Public License.

Notwithstanding any other provision of this License, if you modify the Program, your modified version must prominently offer all users interacting with it remotely through a computer network (if your version supports such interaction) an opportunity to receive the Corresponding Source of your version by providing access to the Corresponding Source from a network server at no charge, through some standard or customary means of facilitating copying of software. This Corresponding Source shall include the Corresponding Source for any work covered by version 3 of the GNU General Public License that is incorporated pursuant to the following paragraph.

Notwithstanding any other provision of this License, you have permission to link or combine any covered work with a work licensed under version 3 of the GNU General Public License into a single combined work, and to convey the resulting work. The terms of this License will continue to apply to the part which is the covered work, but the work with which it is combined will remain governed by version 3 of the GNU General Public License.

# B    Server Side Public License Section 13

13. Offering the Program as a Service.

If you make the functionality of the Program or a modified version available to third parties as a service, you must make the Service Source Code available via network download to everyone at no charge, under the terms of this License. Making the functionality of the Program or modified version available to third parties as a service includes, without limitation, enabling third parties to interact with the functionality of the Program or modified version remotely through a computer network, offering a service the value of which entirely or primarily derives from the value of the Program or modified version, or offering a service that accomplishes for users the primary purpose of the Program or modified version.

"Service Source Code" means the Corresponding Source for the Program or the modified version, and the Corresponding Source for all programs that you use to make the Program or modified version available as a service, including, without limitation, management software, user interfaces, application program interfaces, automation software, monitoring software, backup software, storage software and hosting software, all such that a user could run an instance of the service using the Service Source Code you make available.

# C  Elastic License 2.0

Elastic License Acceptance

By using the software, you agree to all of the terms and conditions below. Copyright License

The licensor grants you a non-exclusive, royalty-free, worldwide, non-sublicensable, non-transferable license to use, copy, distribute, make available, and prepare derivative works of the software, in each case subject to the limitations and conditions below. Limitations

You may not provide the software to third parties as a hosted or managed service, where the service provides users with access to any substantial set of the features or functionality of the software.

You may not move, change, disable, or circumvent the license key functionality in the software, and you may not remove or obscure any functionality in the software that is protected by the license key.

You may not alter, remove, or obscure any licensing, copyright, or other notices of the licensor in the software. Any use of the licensor's trademarks is subject to applicable law. Patents

The licensor grants you a license, under any patent claims the licensor can license, or becomes able to license, to make, have made, use, sell, offer for sale, import and have imported the software, in each case subject to the limitations and conditions in this license. This license does not cover any patent claims that you cause to be infringed by modifications or additions to the software. If you or your company make any written claim that the software infringes or contributes to infringement of any patent, your patent license for the software granted under these terms ends immediately. If your company makes such a claim, your patent license ends immediately for work on behalf of your company. Notices

You must ensure that anyone who gets a copy of any part of the software from you also gets a copy of these terms.

If you modify the software, you must include in any modified copies of the software prominent notices stating that you have modified the software. No Other Rights

These terms do not imply any licenses other than those expressly granted in these terms. Termination

If you use the software in violation of these terms, such use is not licensed, and your licenses will automatically terminate. If the licensor provides you with a notice of your violation, and you cease all violation of this license no later than 30 days after you receive that notice, your licenses will be reinstated retroactively. However, if you violate these terms after such reinstatement, any additional violation of these terms will cause your licenses to terminate automatically and permanently. No Liability

As far as the law allows, the software comes as is, without any warranty or condition, and the licensor will not be liable to you for any damages arising out of these terms or the use or nature of the software, under any kind of legal claim.

Definitions

The licensor is the entity offering these terms, and the software is the software the licensor makes available under these terms, including any portion of it.

you refers to the individual or entity agreeing to these terms.

your company is any legal entity, sole proprietorship, or other kind of organization that you work for, plus all organizations that have control over, are under the control of, or are under common control with that organization. control means ownership of substantially all the assets of an entity, or the power to direct its management and policies by vote, contract, or otherwise. Control can be direct or indirect.

your licenses are all the licenses granted to you for the software under these terms.

use means anything you do with the software requiring one of your licenses.

trademark means trademarks, service marks, and similar rights.