

WHITE PAPER

2021 Open Source Database Trend Report

Executive Summary

In this report, we map the trajectory of top open source data technologies — based on two surveys: The first, a public survey of over 200 development professionals, and the second, a survey of our team of Enterprise Architects.

Contents

Executive Summary 1

Understanding the Open Source Database Landscape 3

Public Survey Results 3

 Job Title.....3

 Company Size4

 Database Survey Findings.....4

Internal Survey Results 5

Final Thoughts 7

Understanding the Open Source Database Landscape

Today, the most popular options available for data engines and data management are open source. But choosing the right open source data technology — one that simultaneously fulfills the needs of the organization and has a high likelihood of remaining relevant and stable for years to come — requires a good understanding of the open source data technologies landscape. That means understanding the trajectory of individual open source technologies and how they fit into larger needs and trends within the market is a critical component of any IT infrastructure plan.

The two surveys presented in this report were designed to give two unique viewpoints into the data technology landscape (including relational database management systems, NoSQL databases and data engines). The first, a public survey, was designed to capture data related to the trajectory of open source databases. The second, a survey of our team of Enterprise Architects, was designed to showcase our expert opinions of the databases we feel are important within the enterprise landscape.

In subsequent sections of the paper, we look at the results of the two surveys, and provide analysis on the results from our team of open source experts.

Public Survey Results

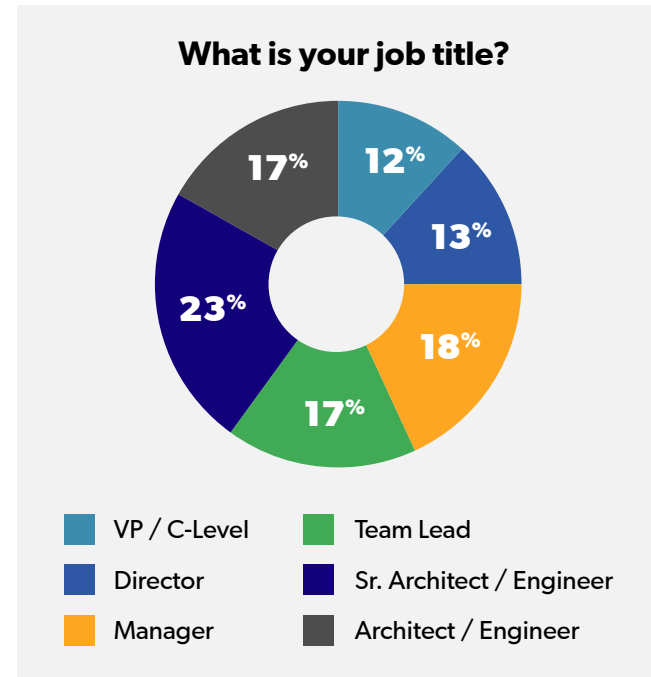
For part one of this report, we conducted a public survey of development professionals. The survey, which ran between May and June 2021, received a total of 211 responses.

The survey was centered on database technologies, with respondents asked to share their knowledge, experiences, and plans for working with top open source options.

To help frame the results, we asked respondents to share their job title and company size.

JOB TITLE

In our first question, we asked respondents to pick the job title that most closely matched their own.



Respondents leaned technical, with technical roles (Architect / Engineer, Senior Architect / Engineer, and Team Lead) constituting 57% of respondents, and management roles (Manager, Director, and VP / C-Level) making up the remaining 43%.

Among individual titles, Senior Architect / Engineer was the leading job title at 23%, with Manager roles at 18%, Team Lead at 17%, Architect / Engineer at 17%, Director at 13%, and VP / C-Level at 12%.

COMPANY SIZE

Next, we asked respondents to share the approximate size of the company they work for.

Respondents to the survey skewed larger, with companies over 100 employees representing 61% of respondents.

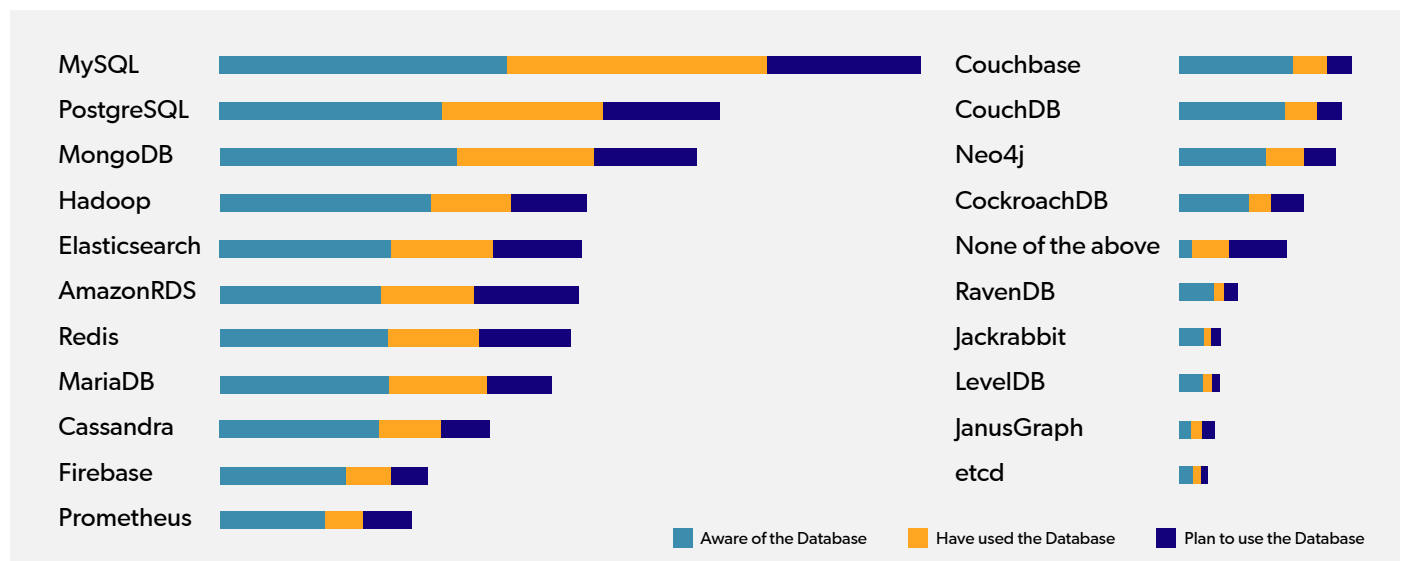
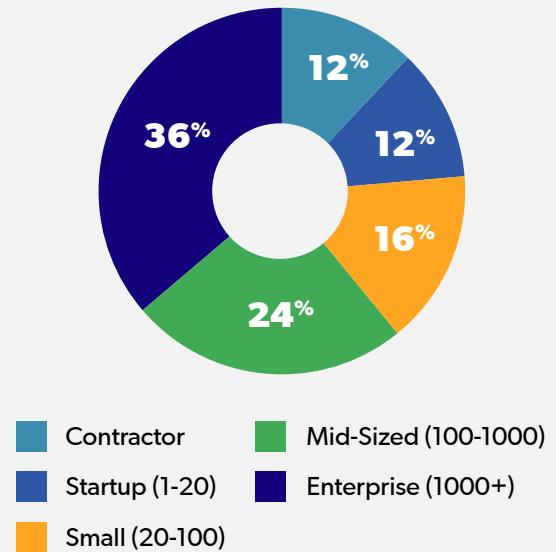
By company size, 36% of respondents work at Enterprise companies, 25% at mid-sized companies, 16% at small companies, and 12% at startup companies. The remaining 12% reported working as contractors or freelancers.

DATABASE SURVEY FINDINGS

For the survey, we asked respondents to share their knowledge, use, and plans for use of 20 popular open source database technologies.

The data, as presented below, combines the responses to provide a sense of the overall importance of the database in current and planned development.

What is the size of your company?



MySQL was the most known, most used, and had the most respondents planning on using it within the next 12 months.

While PostgreSQL surprisingly trailed MongoDB in awareness, it was more used and had more respondents planning on use within the following 12 months.

Rounding out the top five, Hadoop and Elasticsearch were similar in total score, but Hadoop was more well known to respondents, while Elasticsearch had the edge on use and planned use.

TOP RATIONAL DATABASE



TOP GRAPH DATABASE



TOP WIDE COLUMN DATABASE



TOP KEY VALUE DATABASE



TOP DOCUMENT DATABASE



Vince Cox

ENTERPRISE ARCHITECT AT OPENLOGIC

"There are a few surprises here, but one of the big ones is Cassandra being ranked in the middle of the pack. However, when you look at the number of respondents from enterprise organizations it makes a bit more sense. Larger enterprises often make large, ongoing investments into their data infrastructure technologies. Making a change at that scale needs to present a clear (often monetary) benefit. Otherwise, it's just adding risk without a reward.

Just look at Facebook, who, despite being an apparent fit for Cassandra, have stuck with MySQL. Even for changes to MySQL, they've only upgraded their MySQL version when it benefits them directly (e.g. their move from MySQL 5.6 to 5.8, instead of 5.6 to 5.7 to 5.8)."

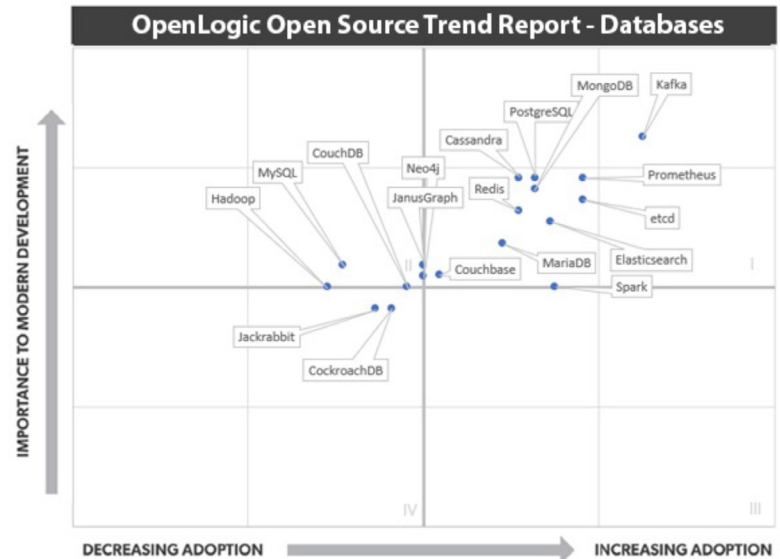
Internal Survey Results

For this segment of the report, we asked our team of Enterprise Architects to rate databases based on two criteria:

1. On a scale of 1-5, with 1 meaning a rapid decrease in adoption, and 5 representing a rapid increase in adoption.
2. On a scale of 1-5, with 1 meaning the technology is unimportant to modern development, and 5 meaning the technology is very important to modern development.

After the results were gathered, we plotted the average of the collective responses to the quadrant chart below.

While we wanted to focus primarily on databases, there are a few technologies listed here that were part of our broader previous trend report on data technologies.



Quadrant #1

Technologies listed in quadrant 1 were seen by our team as important to modern development, and increasing in adoption. These technologies (in order of average rating) include: Kafka, Prometheus, PostgreSQL, MongoDB, etcd, Redis, Elasticsearch, MariaDB, Couchbase, Spark, Neo4j, and JanusGraph.

That said, it's important to note that Spark, Neo4j, and JanusGraph were borderline inclusions. Spark was seen as increasing in adoption, but only as moderately important to modern development. Neo4j and JanusGraph, on the other side, were seen as slightly more important to modern development, but experiencing no change in adoption.

Quadrant #2

Technologies listed in quadrant 2 were seen by our team as decreasing in adoption, but still important to modern development. These technologies (in order of rating) include CouchDB, MySQL, and Hadoop.

Among these, MySQL was listed as the most important technology for modern development, while CouchDB had the least negative sentiment regarding adoption. CouchDB and Hadoop bordered on quadrant 4, meaning our team found them only moderately important to modern development, but decreasing in adoption.

Quadrant #3

Our survey had no technologies listed in quadrant 3, meaning our team saw no listed technologies that were both increasing in adoption, but unimportant to modern development.

Quadrant #4

Technologies listed in quadrant 4 were seen by our team as both unimportant to modern development, and decreasing in adoption.

The two technologies in quadrant 4 (in order of average rating) were CockroachDB and Jackrabbit.



Javier Perez
CHIEF OSS EVANGELIST
PERFORCE SOFTWARE

"The first thing I noticed in looking at our internal survey results is that the data technologies, like Cassandra, Kafka, Spark, Elasticsearch, and PostgreSQL are seeing increasing adoption. Data and AI technologies are driving the adoption of data-driven business decisions and automating processes with a level of sophistication not available just a few years ago.

A few other general notes:

- Elasticsearch was likely underrated by our team because we work with the open source project version and not the commercial edition that is popular within the enterprise.*
- Hadoop is still used extensively, but I think our internal survey is right to recognize it as declining in adoption.*
- Surprised to see CockroachDB rated as a technology decreasing in adoption. That placement might reflect our experience in working with a segmented market.*
- Generally, the internal survey results do a good job of showcasing the popularity of different types of open source data technologies."*

Final Thoughts

It's hard to think of an area of open source more integral to the digital economy than databases and data technologies. With the gross amount of data that enterprises create, ingest, and process, the underlying technologies are inarguably critical to business success. Whether that's the data that drives business decisions, or reflects the state of the business, choosing and maintaining the right database technologies is key.

Still, the same hard truth that applies to open source data technologies also applies for any open source technology: Choosing a technology only based on trend is a poor choice, and one that can have lasting consequences. Choosing the right technology should always be based on business needs and compatibility between technology stacks – with one eye on the trajectory and longevity of that technology, and the other on how you plan to support and optimize it going forward.

While it was not discussed in the surveys, the ecosystem of open source software around the selected data technology is also an important factor. Like any other software, users should consider the health of the user community, availability of complementary software, and internal expertise for the technology.

We hope this guide served as a good conversation starter regarding your current and planned data infrastructure. If you have any questions about the results, or just want to chat regarding your plans, please contact us today.

About Perforce

Perforce powers innovation at unrivaled scale. Perforce solutions future-proof competitive advantage by driving quality, security, compliance, collaboration, and speed – across the technology lifecycle. We bring deep domain and vertical expertise to every customer, so nothing stands in the way of success. Privately held and funded by Clearlake Capital and Francisco Partners, our global footprint spans more than 80 countries and includes over 75% of the Fortune 100. Perforce is trusted by the world's leading brands to deliver solutions to even the toughest challenges. Accelerate technology delivery, with no shortcuts. [Get the Power of Perforce.](#)

Get Support for Your Open Source

Whether you're planning your next integration, or need support for your existing data infrastructure, OpenLogic can help support your goals.

Talk to an expert today to see how we can help.

TALK TO AN EXPERT

<https://www.openlogic.com/talk-to-expert>