



Hanoi, Vietnam 2024

**PERCONA  
UNIVERSITY**

**Open Source Databases Meetup**

In partnership with





# Welcome



Peter Zaitsev,  
Founder at Percona  
November 2, 2024

In partnership with



Hanoi, Vietnam 2024

**PERCONA**  
UNIVERSITY

Open Source Databases Meetup



Thank you  
to our Partner



Hanoi, Vietnam 2024

**PERCONA**  
**UNIVERSITY**

Open Source Databases Meetup

**Sameer Kumar**

Your friendly  
Percona Sales Person

sameer.kumar@percona.com

LinkedIn: -sameer-kumar-





# Bring Percona University to Your City!



# WE ARE HIRING

We are **hiring**.  
Check our openings.



Current openings include:

- Senior Software Engineer (PostgreSQL)
- Support Engineer (PostgreSQL)
- PostgreSQL Evangelist
- ...and more!





# State of Open Source Databases



Hanoi, Vietnam 2024

**PERCONA**  
UNIVERSITY

Open Source Databases Meetup

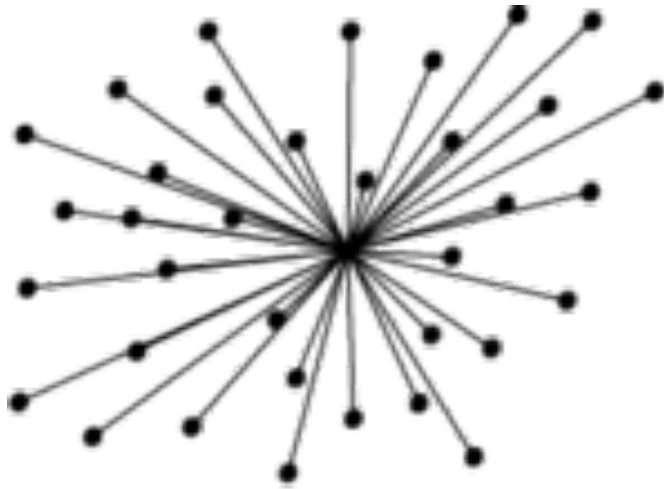
In partnership with



A glowing lightbulb is centered in the frame, set against a background of a sunset or sunrise with a gradient of orange, red, and blue. The lightbulb is illuminated from within, casting a warm glow. The word "Innovation" is written in a large, white, sans-serif font across the middle of the image, partially overlapping the lightbulb. The overall mood is one of inspiration and creative thinking.

# Innovation

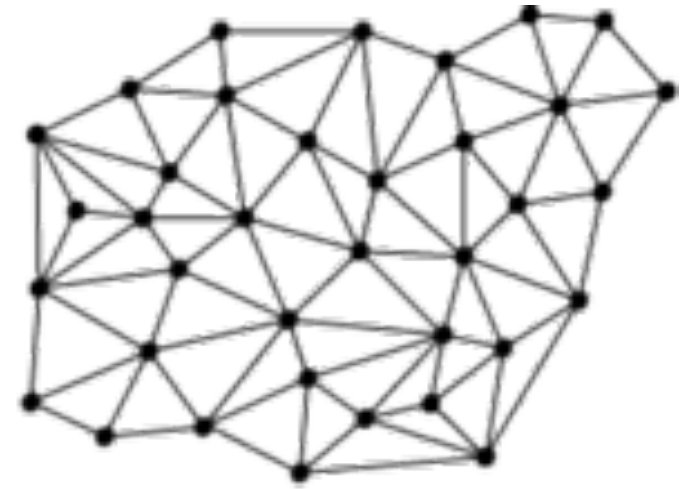




centralised



decentralised



distributed

# Distributed Databases




# Separation of Storage and Compute



# Serverless



A LEGO-style winged unicorn with a rider on a beach at sunset. The unicorn is light purple with large, translucent pink wings. It has a yellow mane and tail with blue and pink floral patterns. A rider with blonde hair, wearing a green and yellow outfit and a crown, sits on its back. The background shows a beach, waves, and a large tree trunk under a sunset sky with stars.

# HTAP – Hybrid Transactional Analytical Processing





# Data Pipelines

# Multiple Databases

**Microservices**

**Developers Authority**

**Cloud**





**Not just Relational  
Any More**



**Time Series**



**Graph**



**Data Structures**

# Different Models



# Multi-Model

Some Databases Support Multiple Data Models, Some even Talk Different Languages/Protocols





# Vector Search

**Technology Supporting building  
applications which use AI**



**AI Algorithms do not work with words and Images inside**



**They work with Vectors**



**Objects encoded to Vector form called "Embeddings"**



**Finding Similar Vectors is a key task**



**Traditional Database Algorithms do not handle it well**

# What's Up with Vectors

<https://manticoresearch.com/blog/vector-search-in-databases/>



## Specific Tasks



Pattern Recognition



Semantic Search



RAG – “Retrieval Augmented Generation”

# Vector search support in databases

## Opensource vector dbs

Milvus	2019
Vespa	2020
Weaviate	2021
Qdrant	2022

## Opensource dbs and search engines

PostgreSQL	2021
Lucene	2021
Opensearch	2022
Redis	2022
SOLR	2022
Cassandra	<b>2023</b>
Typesense	<b>2023</b>
Clickhouse	<b>2023</b>
Manticore Search	<b>2023</b>
Meilisearch	<b>2023</b>
MariaDB	<b>In progress</b>
MySQL	Not yet

## Non-open source dbs

Elasticsearch	2019
Oracle	<b>2023</b>
MongoDB	<b>2023</b>

## Clouds

Pinecone	2019
Amazon Elasticsearch / Opensearch	2020
Google Cloud Platform	2021
Alibaba Cloud AnalyticDB	<b>2023</b>
Azure	<b>2023</b>
Amazon DocumentDB	<b>2023</b>
Cloudflare Vectorize	<b>2023</b>



# What Is the Biggest Factor Impacting Open Source Now?

Many will Say:  
**Cloud**





**Maximize and  
Simplify  
Adoption**

**Change  
Opportunities  
for Monetization**

# Cloud Impact

A close-up portrait of Marten Mickos, a man with short brown hair and blue eyes, wearing a white shirt. The background is a solid light gray.

# Remember...

**Marten Mickos: "Open  
Source Is Not a Business  
Model"**

# Open Source Ownership and Governance

- Foundation Driven (Multiple Vendors)
- Single Vendor Driven





**CLOUD HELPS TO  
ACCELERATE ADOPTION**



**CLOUD CHANGES WHO  
CAPTURES THE VALUE**

# Foundation Based Open Source



# Single Vendor

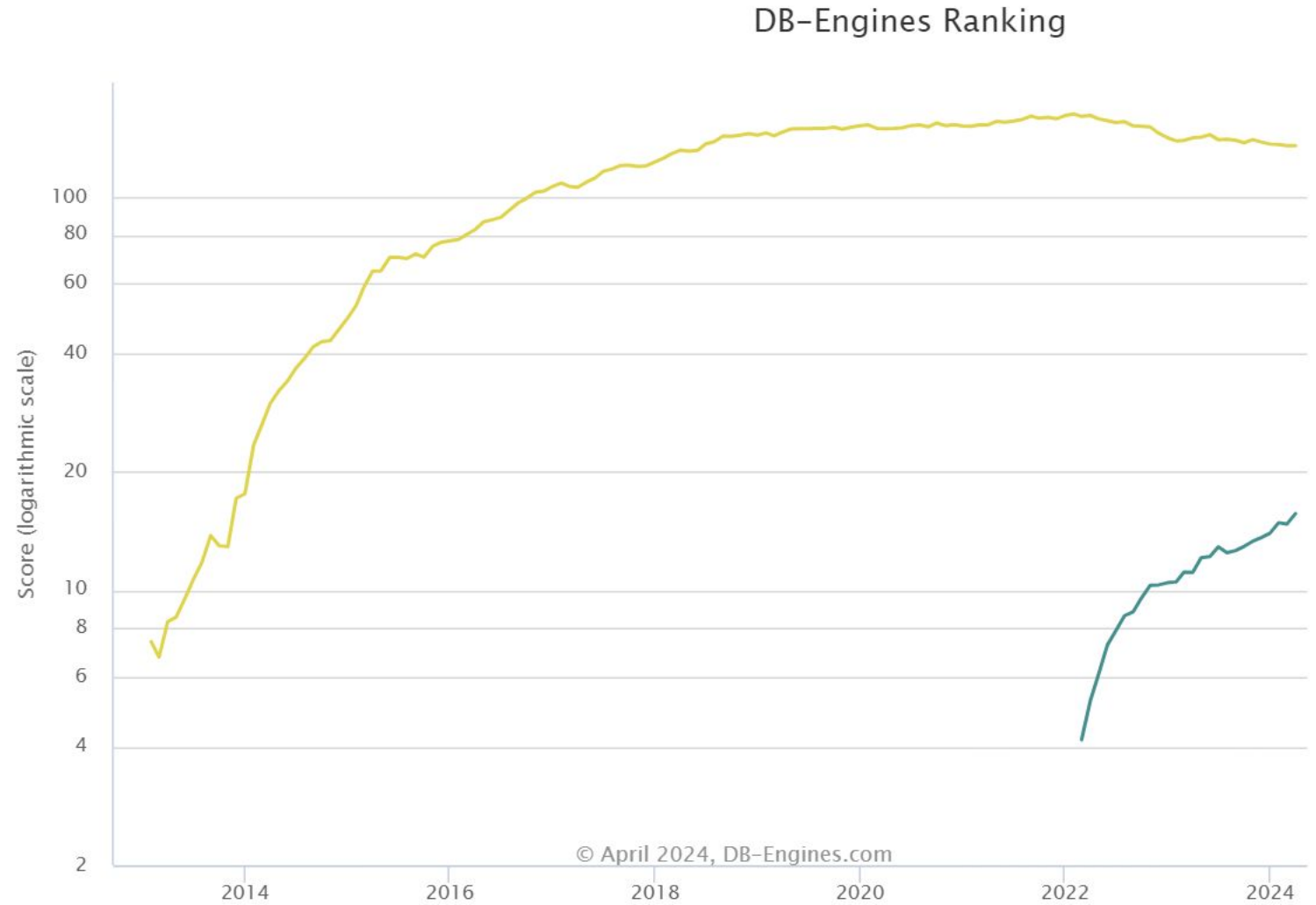
- Tend to Be Venture Funded or Public Companies
- Feared of Competition with Cloud Vendors

# Fully or Partially Abandoning Open Source Licenses

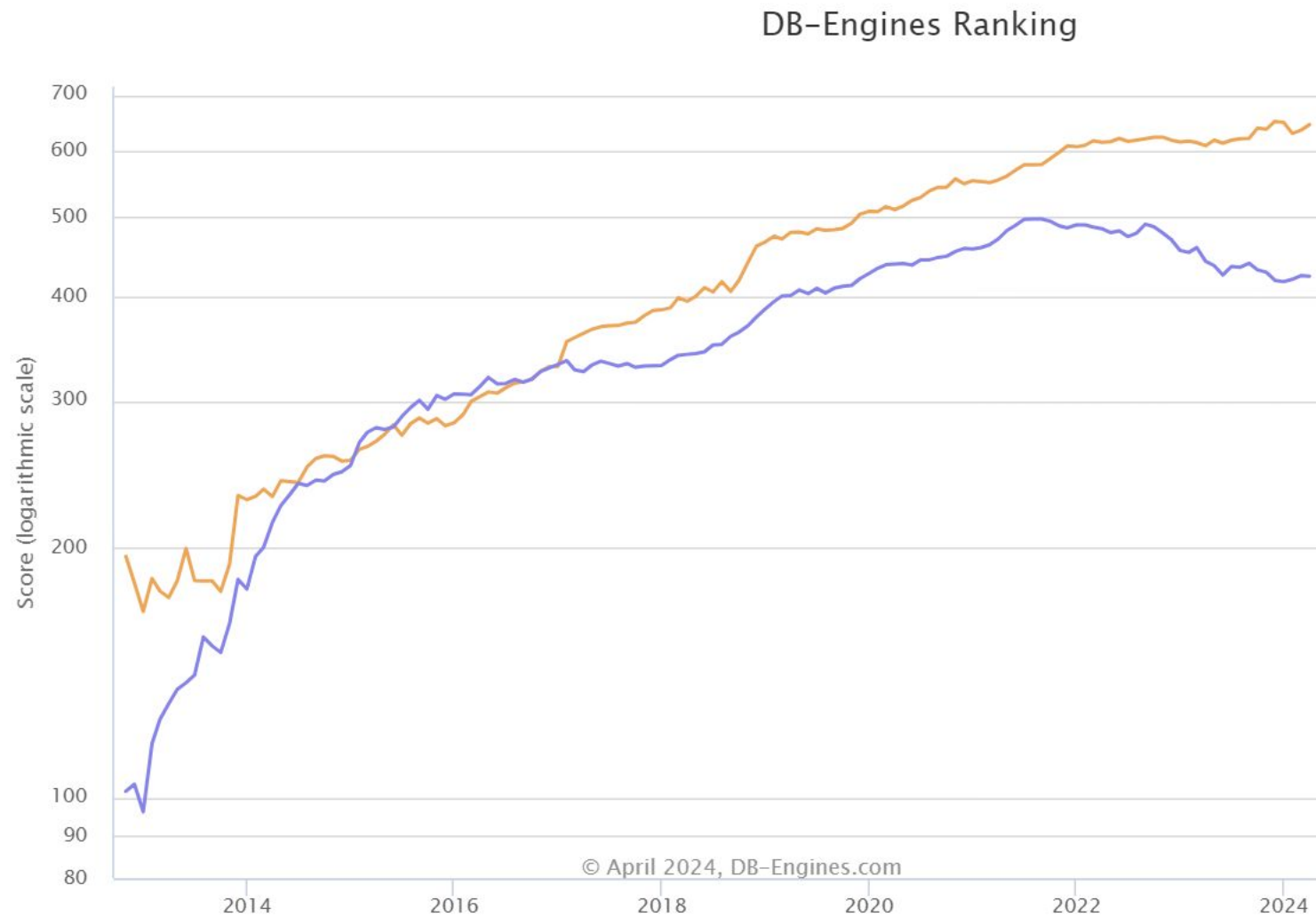




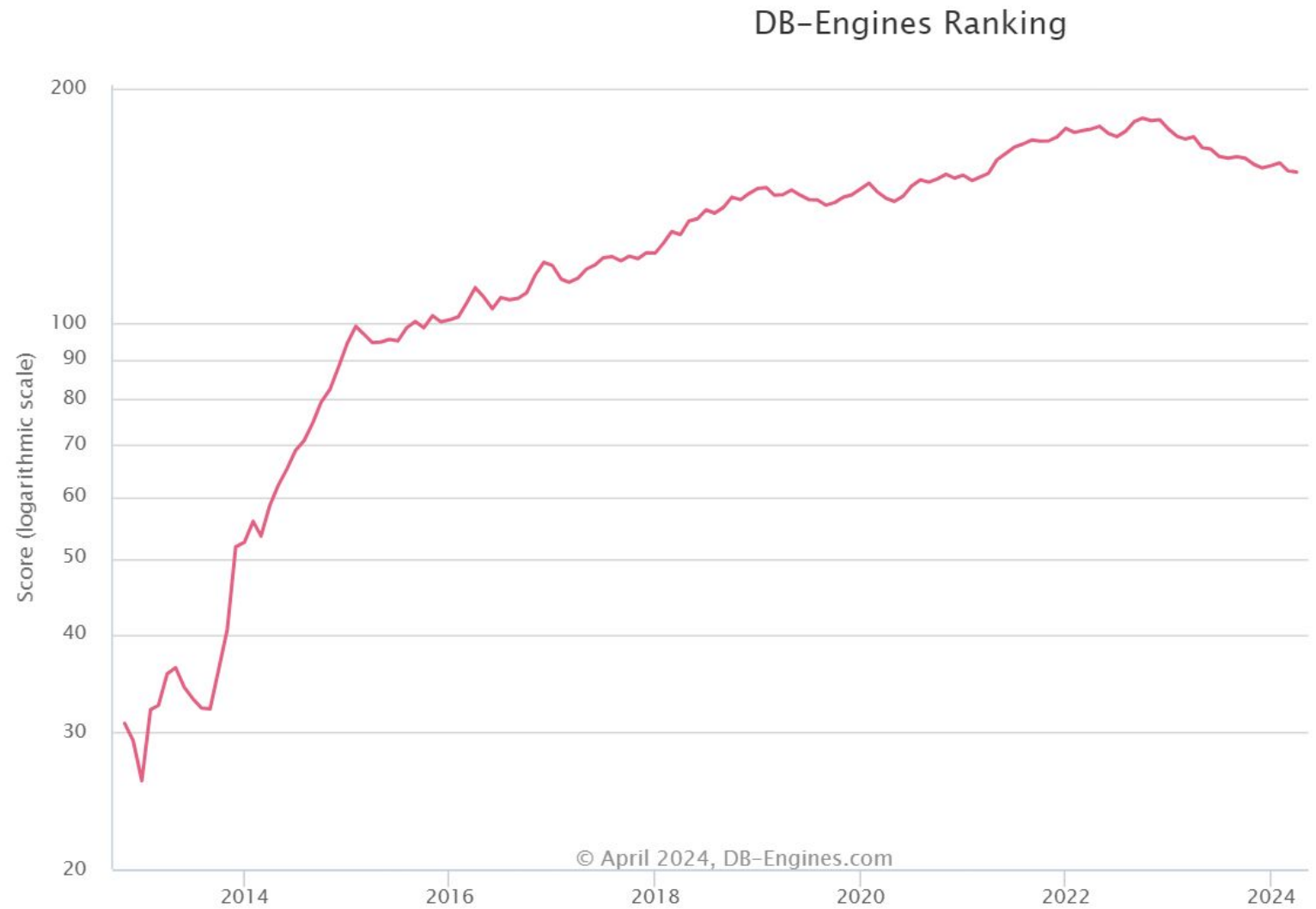
# Elastic vs OpenSearch after Licence Change



# MongoDB vs PostgreSQL



# Redis





# Linux Foundation is Stepping Up!

 SIGN IN / UP

The  Register®

DATABASES

## Linux Foundation marshals support for open source alternative to Redis

7 

Follows the vendor's decision to overhaul licensing of the popular cache database

 [Lindsay Clark](#)

Wed 3 Apr 2024 // 11:15 UTC



Cloud giants AWS, Google, and Oracle have come out in support of a Linux Foundation open source fork of Redis, the popular in-memory database frequently used as a cache, following changes to its licensing.

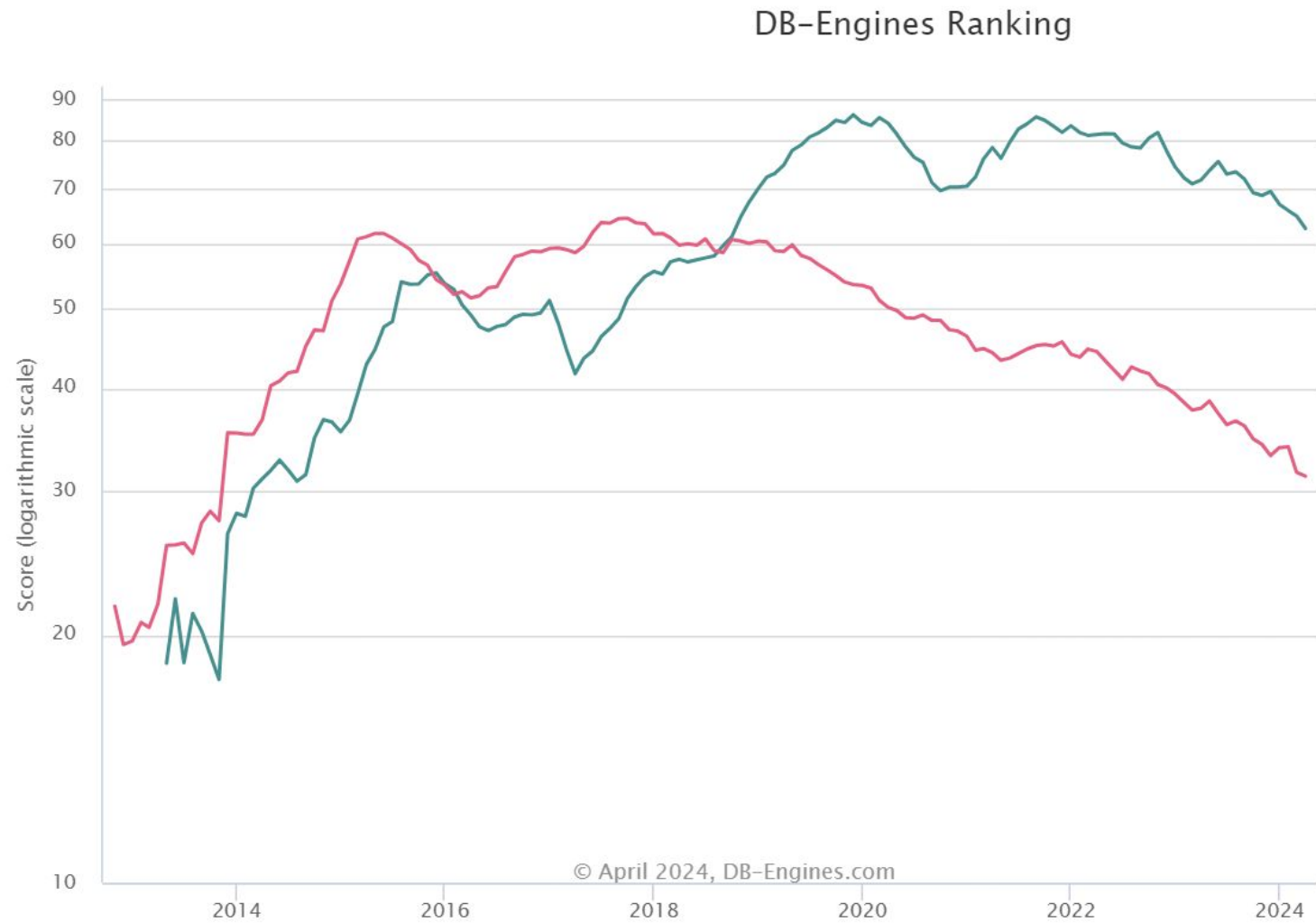
Last month, [Redis confirmed it was shifting its main key-value store system to a dual-license approach](#), imposing far more restrictive terms. Previously, the source code was available under the Berkeley Software Distribution (BSD) 3-clause license, which allows developers to make commercial use of the code without paying.

[https://www.theregister.com/2024/04/03/open\\_source\\_redis\\_alternative/](https://www.theregister.com/2024/04/03/open_source_redis_alternative/)

# Percona Joins the Effort!



# Hive (Blue) and HBase







# Primary Goal of the License Change?

- **Creating Monopoly on DBaaS Market**



**STATE OF ART SIMPLICITY**



**HIGH LEVEL OF  
AUTOMATION**




**MAXIMIZES DEVELOPERS  
FOCUS ON APPLICATION**

# Why DBaaS?

# What Is the Problem with Monopoly on DBaaS?

MONOPOLY





# No Different from Proprietary Software

Using DBaaS is a very different  
skill compared to rolling your  
own database setup





# DBaaS Vendor Lock-In

It may not be so painful **now**, but it is going to be painful in the **future**



# History Lesson

**Oracle used to Save its customer from IBM  
Hardware Lock-in with Mainframe computers**





# Why Data is Special

**Moving Data is Expensive. Moving Lots of Data is Very Expensive – Principle of Data Gravity Applies**



# Not All Is Lost



# Have We Been Here Before?

## 2000s



## 2020s





# Operating Systems



# Open Source Catches Up Again



- **Lock-in with Cloud Vendor**
- **Use Proprietary Solutions**
- **Highly Differentiated Cloud**



**CLOUD NATIVE  
COMPUTING FOUNDATION**

- **Freedom to Run Anywhere**
- **Use Open Source**
- **Cloud Is Commodity**
- **Customer**
- **Choice of Vendors**



# Giving Cloud Its Originally Intended Role of Commodity Infrastructure

## What is Cloud Computing?

An analogy: think of electricity services...

You simply plug into a vast electrical grid managed by experts to get a low cost, reliable power supply – available to you with much greater efficiency than you could generate on your own.

Power is a utility service - available to you on-demand and you pay only for what you use.



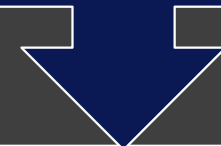
# Kubernetes

# kubernetes

**Kubernetes is universally available**



**Kubernetes is getting better for stateful applications**

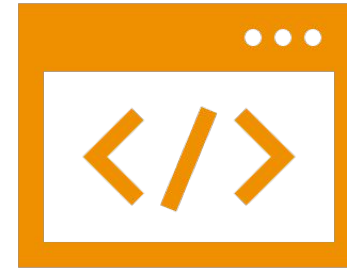


**Kubernetes Operators are available for most popular Open Source Databases**





**Day 1 and Day 2 Automation, Toil  
Reduction Similar to DBaaS**



**UX is Different, Requires  
Kubernetes Expertise**

# What's Up with Kubernetes Operators?

# Can We Build DBaaS on Kubernetes?



kubernetes

Many Modern  
DBaaS are  
Built on  
Kubernetes  
Operators





# Open Source DBaaS Experience with Percona Everest

The screenshot displays the Percona Everest configuration interface, specifically the 'Resources' step. The interface is divided into three main sections: a left sidebar with navigation icons, a central configuration area, and a right sidebar with a 'DATABASE SUMMARY'.

**Left Sidebar:** Contains navigation icons and a progress indicator with six steps. Step 5 is highlighted, indicating the current step.

**Central Configuration Area:**

- Resources:** Configure the resources your new database will have access to.
- Number of nodes:** Three buttons: '1 node', '2 nodes', and '3 nodes'. '3 nodes' is selected.
- Resource size per node:** Four buttons: 'Small', 'Medium', 'Large', and 'Custom'. 'Custom' is selected.
- CPU:** A box containing '3' and 'CPU', followed by 'x 3 nodes = 9 CPU'. Below it, 'Estimated available: 12.142 CPU'.
- MEMORY:** A box containing '8' and 'GB', followed by 'x 3 nodes = 24 GB'. Below it, 'Estimated available: 62.44959552 GB'.
- DISK:** A box containing '250' and 'GB', followed by 'x 3 nodes = 750 GB'.
- Navigation:** A blue 'Continue' button and a grey 'Cancel' button.

**Right Sidebar: DATABASE SUMMARY**

- 1. Basic Information** (with an edit icon):
  - Namespace: app
  - Type: PostgreSQL
  - Name: postgresql-dev
  - Version: 16.1
  - Storage class: standard-rwo
- 2. Resources** (highlighted):
  - Number of nodes: 3
  - CPU: 9 CPU
  - Memory: 24 GB
  - Disk: 750 GB
- 3. Backups**
- 4. Point-in-time Recovery**
- 5. Advanced Configurations**
- 6. Monitoring**

**Bottom of the window:** A blue 'Continue' button, a grey 'Cancel' button, and a status indicator for '6. Monitoring Enabled' with an edit icon.

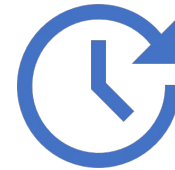
# Summary



**Fantastic Pace  
Innovation**



**Evolving Open  
Source Models**



**Great Future**



# Thank you! Let's Connect!

<https://www.linkedin.com/in/peterzaitsev/>

<https://twitter.com/PeterZaitsev>

<http://www.peterzaitsev.com>

In partnership with

