



/ANITA  
B.ORG 2024  
**GRACE HOPPER  
CELEBRATION**

me  
+we



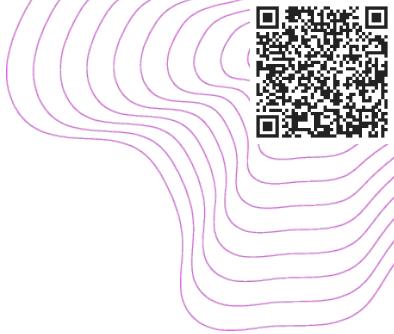
# Milvus: An Open-Source Vector Database

*Learn about Deep  
Learning  
embeddings*

*Create a Vector  
Database*

*Build cool stuff!*

me  
+we



# Welcome!



# Stefan Webb

Developer Advocate, Zilliz

[stefan.webb@zilliz.com](mailto:stefan.webb@zilliz.com)

<https://www.linkedin.com/in/stefan-webb>

[https://x.com/stefan\\_webb](https://x.com/stefan_webb)





## Tim Spann

Principal Developer  
Advocate, Zilliz

[tim.spann@zilliz.com](mailto:tim.spann@zilliz.com)

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

<https://x.com/PaaSDev>



## Jiang Chen

Head of Ecosystems and  
Developer Relations, Zilliz

[jiang.chen@zilliz.com](mailto:jiang.chen@zilliz.com)

<https://www.linkedin.com/in/jiangc1010>

<https://x.com/jiangc1010>



# Today's Schedule

## Morning Session (PT)

9.45 – 10.15	<i>Intro to Milvus and Vector Databases</i>
10.15 – 10.35	<i>Getting Started with Milvus</i>
10.35 – 11.45	<i>Contributing to Milvus; or, Milvus Workshop</i>

## Afternoon Session (PT)

13.15 – 13.45	<i>Intro to Milvus and Vector Databases</i>
13.45 – 14.05	<i>Getting Started with Milvus</i>
14.00 – 15.15	<i>Contributing to Milvus; or, Milvus Workshop</i>



# Searching the Web with Gen AI

Where knowledge begins

How does a vector database work? in

≡ Focus   + Attach

Pro



What is Perplexity AI?



Success rate of Lasik surgery



When will the next iPhone be released?



Vegetables currently in season



# How does a vector database work?

## Sources

What is a Vector Database? - Elastic

elastic · 1

What is a Vector Database & How Does it Work? Use Cases + ...

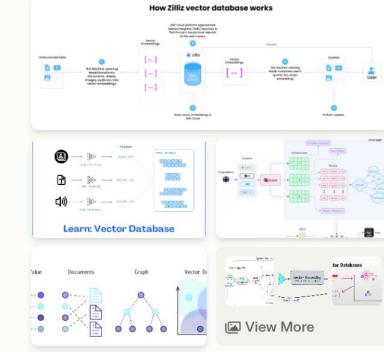
pinecone · 2

What are Vector Databases and How Do They Work? - Zilliz blog

zilliz · 3



View 4 more



Learn: Vector Database

Value Documents Graph

View More

## Search Videos



## Generate Image



## Perplexity

A vector database works by storing, indexing, and searching data represented as high-dimensional vectors. Here's an overview of how vector databases function:

### Vector Representation

Vector databases store data as vector embeddings - numerical representations of data objects like text, images, or audio <sup>1</sup>. These embeddings are created using machine learning models that capture semantic relationships between data points <sup>5</sup>.

### Indexing

Vector databases use specialized indexing algorithms to organize the vector data for efficient retrieval:

- Algorithms like HNSW (Hierarchical Navigable Small World), LSH (Locality-Sensitive Hashing), or PQ (Product Quantization) are used to index vectors <sup>3</sup>.
- Indexing maps vectors to data structures that enable fast similarity searches <sup>5</sup>.

### Querying

When a query is made:

- The query is converted into a vector embedding using the same model used for indexing <sup>1</sup>.
- The database performs an Approximate Nearest Neighbor (ANN) search to find the

Ask follow-up

Pro ↑



# Why is Semantic Search Difficult?

Q | Apple



Q | Rising dough

Rising Dough ✓

Q | Change car tire



VS.



VS.

Proofing Bread ✗

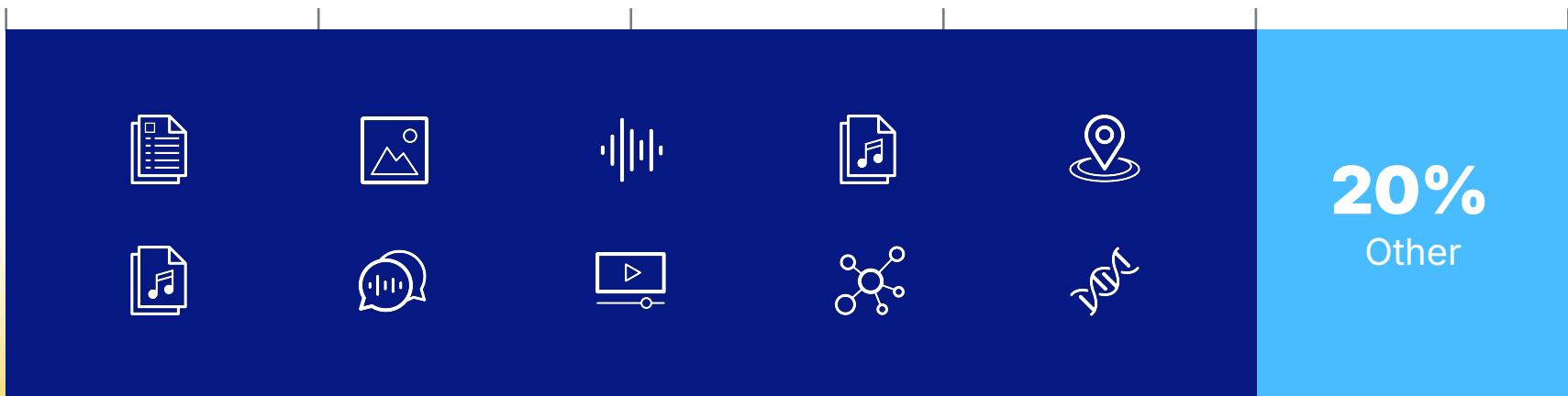
VS.





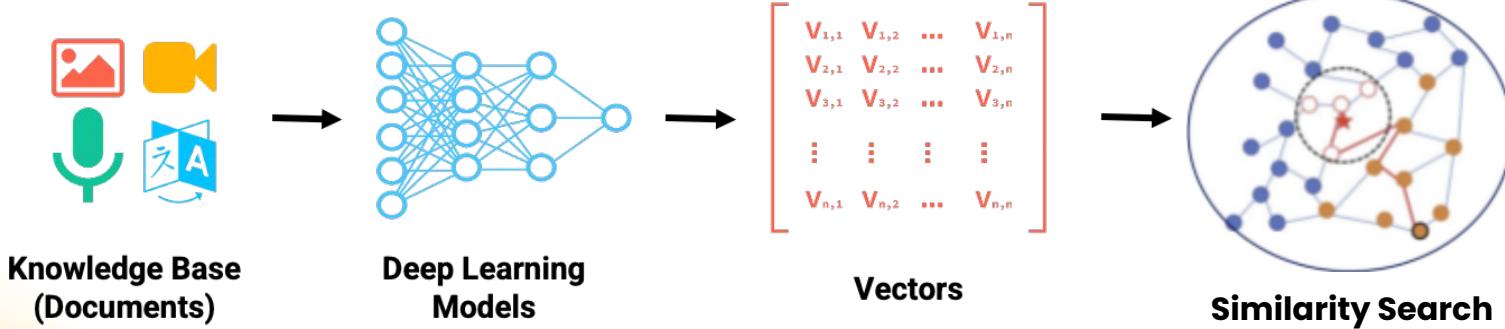
# Why is Semantic Search Important?

80% newly generated data in 2025 will be unstructured data



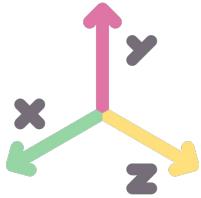


# Solution: Deep Learning



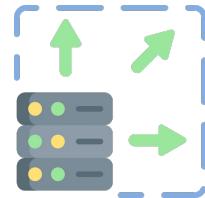


# New Challenge: Search in Vector Spaces



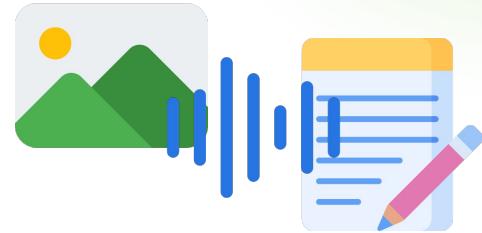
## How to Index and Search?

- High-dimensional
- > 1000 dims



## How to Scale?

- 10–100 million vectors?
- Billions?
- Trillions?
- Billions of users?



## Multiple Data Types?

- Text
- Images
- Audio
- Graphs
- ...



# Milvus: The most widely-adopted vector database

Milvus is an **Open-Source Vector Database** to **store, index, manage, and use** the massive number of **embedding vectors** generated by deep neural networks and LLMs.



400+  
contributors

30K+  
stars

66M+  
docker pulls

2.7K  
+  
forks



# Milvus Users

accenture

airbnb

AT&T

BOSCH

Chegg

CISCO

CISION

COMPASS

Deloitte. ebay

FARFETCH

Grab

IKEA

Inflection

intuit

Microsoft

new relic.

NVIDIA.

OMERS

Olli. Otter.ai

PayPal

paloalto  
NETWORKS

POSHMARK

RUBLOX

salesforce

Shell

shutterstock

T

TREND  
MICRO™

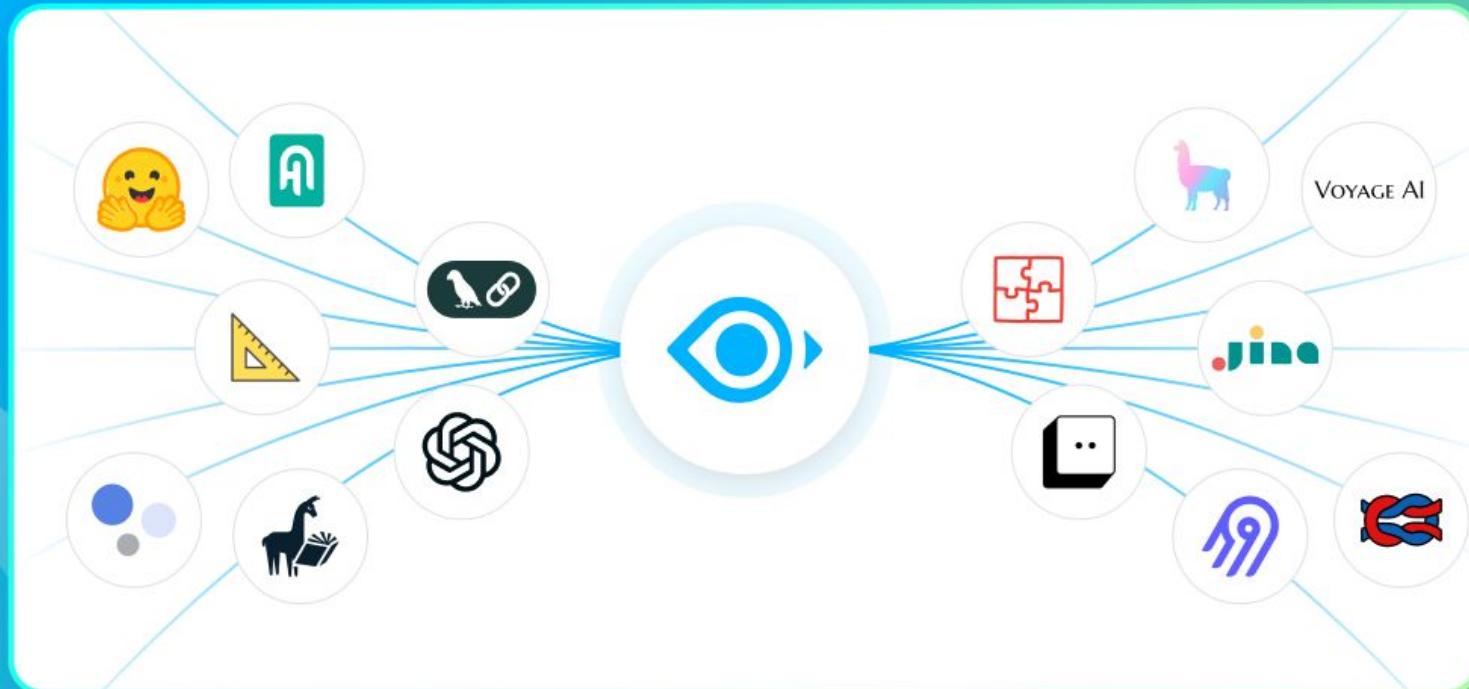
Walmart



ZipRecruiter

zomato

# Seamless integration with all popular AI toolkits





# Why Open-Source?



**Cost-effective**



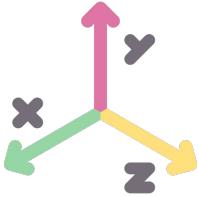
**Innovation**



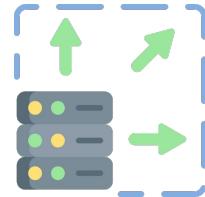
**Community**



# Why Not Traditional Databases?



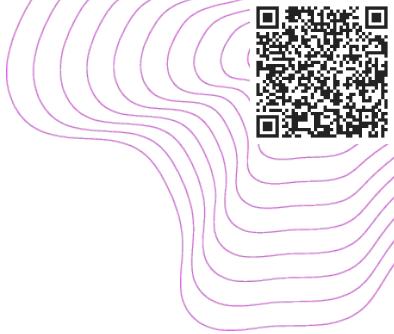
**Suboptimal  
Indexing / Search**



**Scaling**



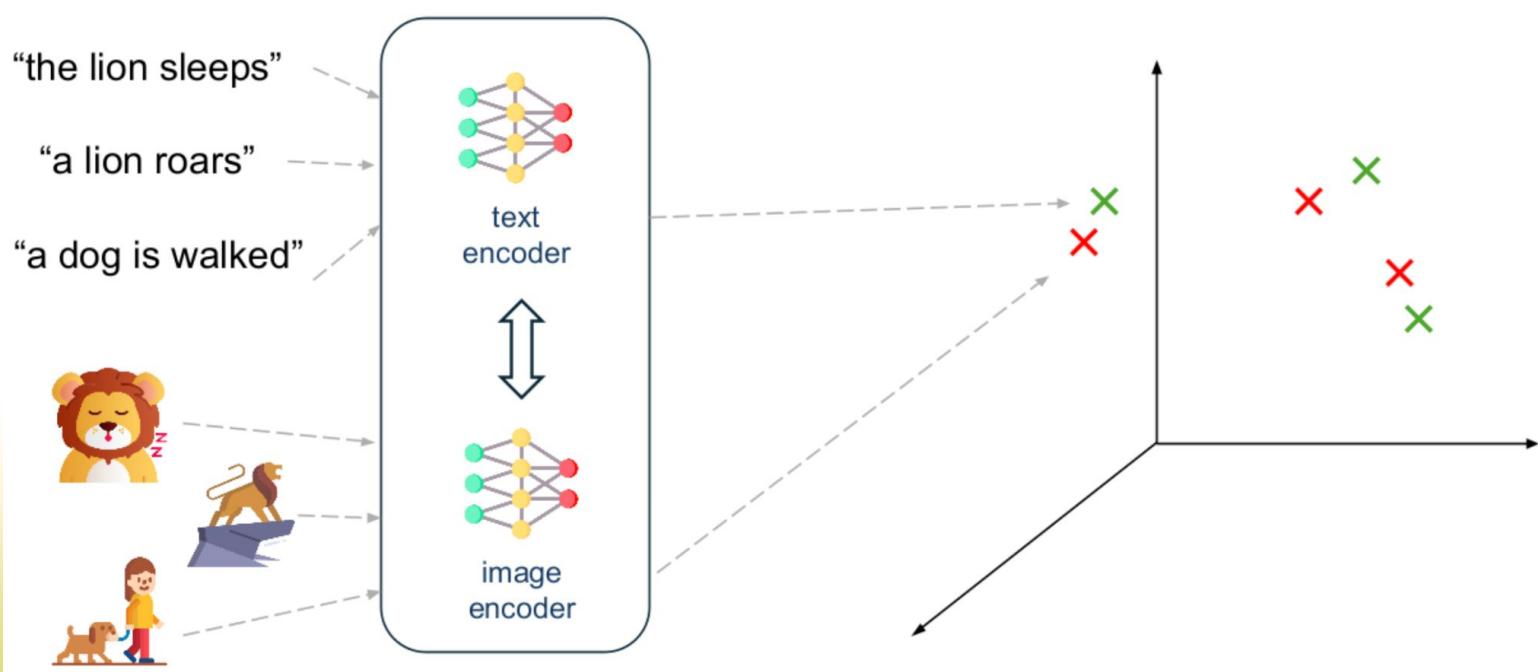
**Inadequate Query  
& Analytics Support**



# All About Vector Databases and Milvus

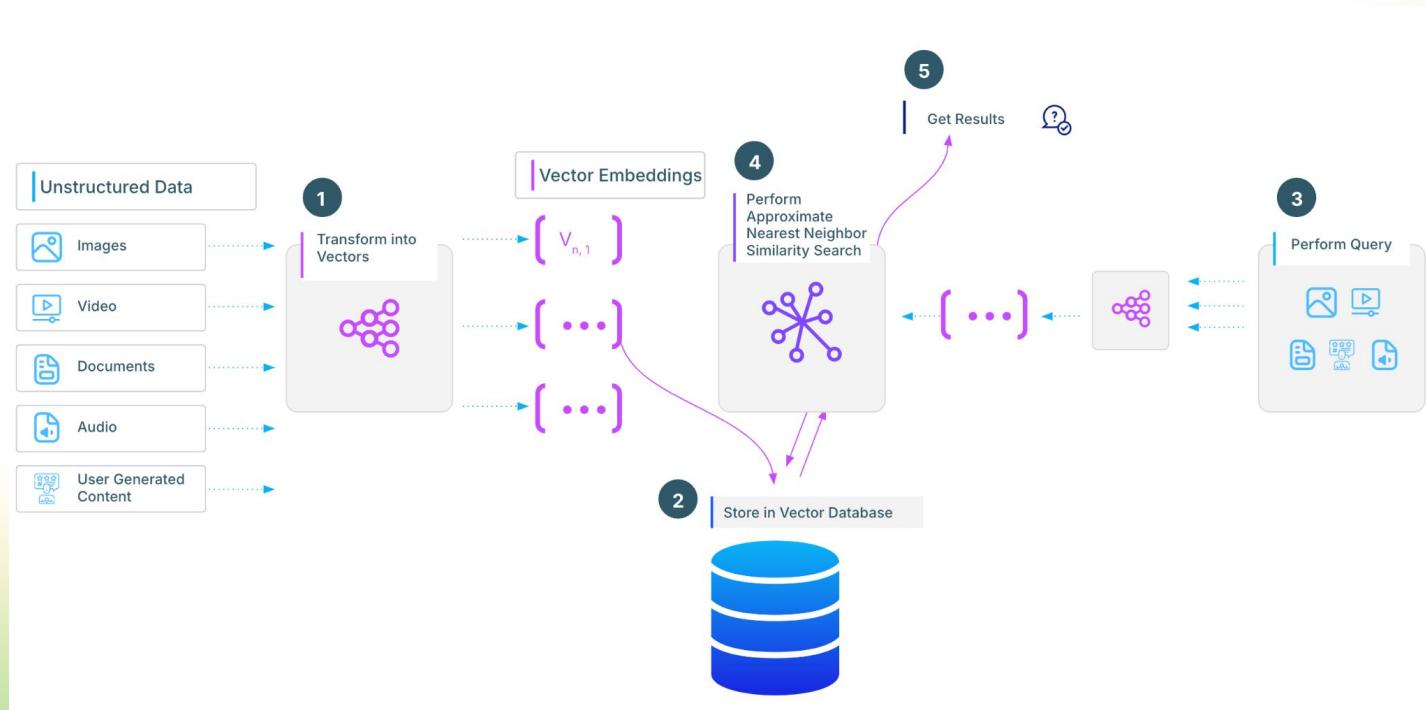


# Semantic Similarity?





# How Does Similarity Search Work?



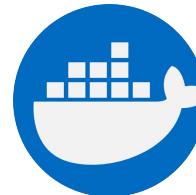


# Deployment Options



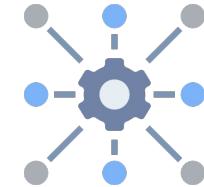
## Milvus Lite

- Locally hosted
- Suitable for prototyping and demos



## Milvus Standalone

- Single remote/local server
- “Medium” scale
- Simplified setup, maintenance, etc. compared to cluster

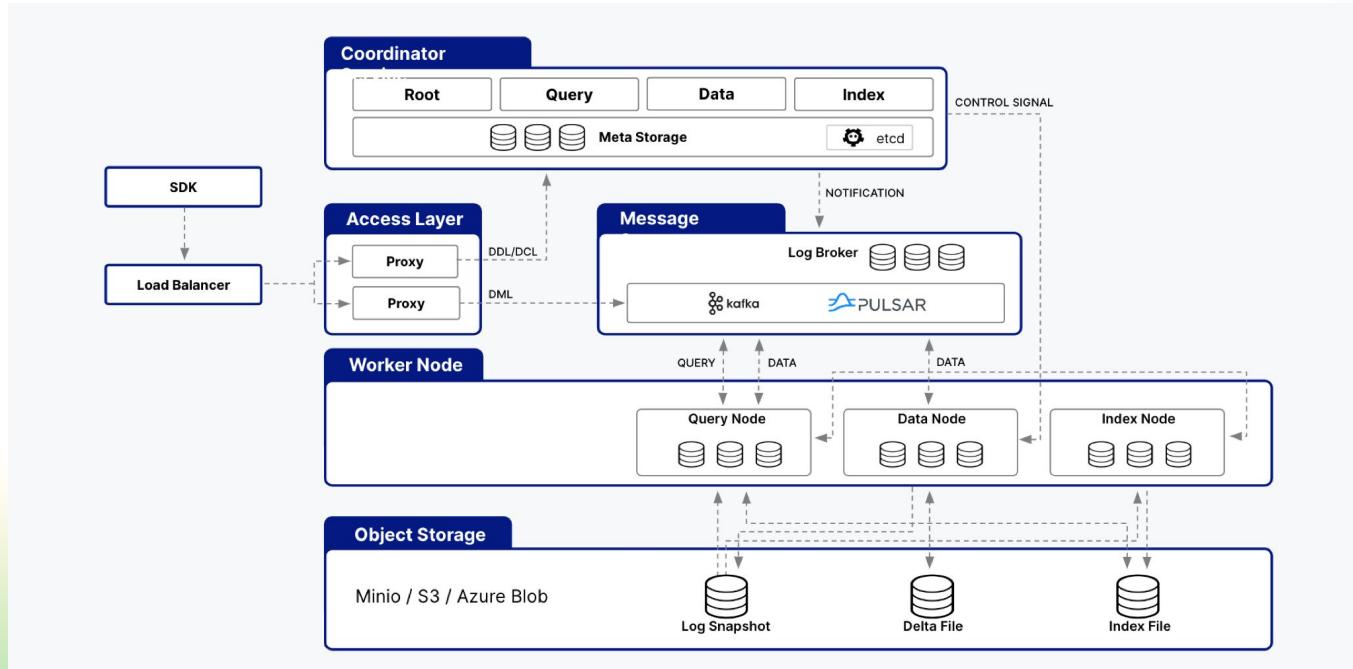


## Milvus Cluster

- Distributed system
- Many different types of nodes
- Scales to 100s of billions of vectors



# Vector Database Architecture





# Benchmarks

## Search Performance Test (10M Dataset, 768 Dim)

Qps (more is better)



## Search Performance Test (10M Dataset, 768 Dim)

Qps (more is better)



## Search Performance Test (5M Dataset, 1536 Dim)

Qps (more is better)



## Search Performance Test (5M Dataset, 1536 Dim)

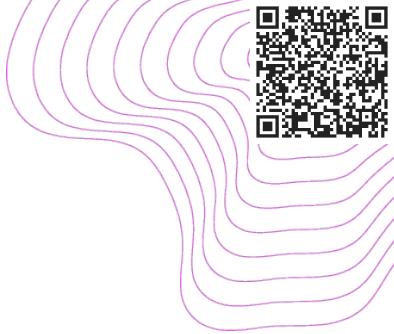
Qps (more is better)



Shows 3-20x faster comparing with open source Milvus

<https://github.com/zilliztech/VectorDBBench>

At least 6x faster than other vector databases



# Demo Time: What Can you Build with Milvus?



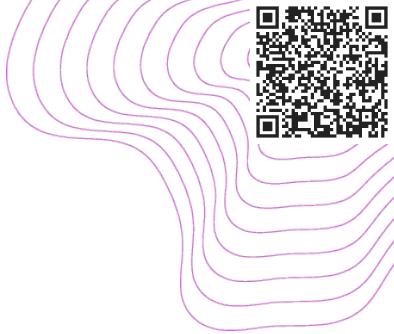
# #1 – Multimodal Image Search

- <https://multimodal-demo.milvus.io/>



## #2 - Chat with Open Source Software

- <https://osschat.io/>



# Getting Started with Milvus and Workshop



# Getting Started

- [Google Colab notebook](#)



# Choose Your Own Adventure





# Contributing Your First Issue

- A few first-time-contributor-suitable issues:
  - [\[Feature Request\]: \[Milvus\] Collection load](#)
  - [\[Feature Request\]: \[Milvus\] Shards Number](#)
  - [\[Feature Request\]: \[Milvus\] Timeout](#)
  - [\[Feature Request\]: \[Milvus\] Partition](#)
- Help add integrations of Milvus and new RAG/agent frameworks:
  - [dynamiq-ai/dynamiq](#)
  - [CrewAI](#)
  - [EmbedAnything](#)
  - [Firebase GenKit](#)
  - [Cognee](#)

<https://zilliz.com/blog/contributing-to-open-source-milvus-beginners-guide>



# Contributing Your First Issue

You can also contribute to Milvus directly:

- [Issues · milvus-io/milvus · GitHub](#)

by following the [contribution instructions](#), and see

- [Contributing to Open Source Milvus: A Beginner's Guide](#)



# Milvus Mini-Projects

*Call for contribution of projects and demos built with Milvus. Get inspired by examples [here](#). We can feature them on Discord.*

Suggestions:

- RAG with Contextual Retrieval
  - [Introducing Contextual Retrieval \ Anthropic](#)
- RAG for an application that requires structured output
  - [GitHub - dottxt-ai/outlines: Structured Text Generation](#)
  - [Structured Outputs - OpenAI API](#)
- RAG on Wikipedia
  - [wikimedia/wikipedia · Datasets at Hugging Face](#)



# Bootcamp Tutorials

## Image Similarity Search



Build a reverse image search system using Milvus paired with YOLOv3 for object detection and ResNet-50 for feature extraction.

[Live Demo](#)

## Question Answering System



Build an intelligent chatbot using Milvus and the BERT model for natural language processing (NLP).

[Live Demo](#)

## Recommender System



Build an AI-powered movie recommendation system using Milvus paired with PaddlePaddle's deep learning framework.

## Video Similarity Search



Build a video similarity search engine using Milvus and a VGG neural network.

## Audio Similarity Search



Build an audio search engine using Milvus paired with PANNs for audio pattern recognition.

## Molecular Similarity Search



Build a molecule similarity search system using Milvus paired with RDkit for cheminformatics.

[Live Demo](#)

## DNA Sequence Classification

Build a DNA sequence classification model using Milvus paired with CountVectorizer.

<https://milvus.io/bootcamp>

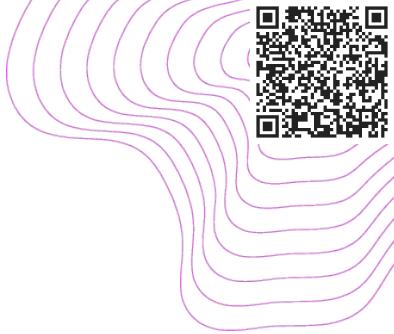


# Workshop: Building Semantic Search with Milvus

- [Google Colab notebook](#)

# Generative AI Resource Hub

- [Generative AI Resource Hub | Zilliz](#)

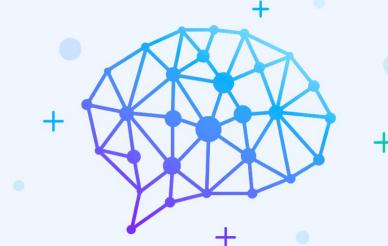


# Open Discussion Time



## Unstructured Data Meetups

Join a Community of Passionate Developers and Engineers  
Dedicated to Gen AI



South Bay  
[RSVP now →](#)



San Francisco  
[RSVP now →](#)



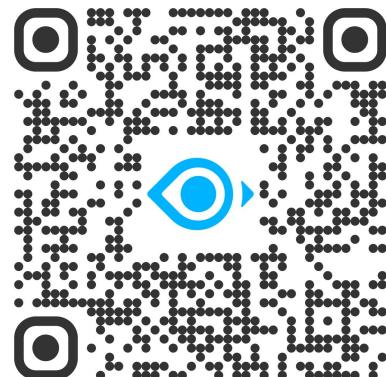
Berlin  
[RSVP now →](#)



New York  
[RSVP now →](#)

*10/8/2024 San Francisco  
10/15/2024 Silicon Valley  
10/23/2024 New York*

Held twice a month





# We're Hiring!

- Engineering Manager, Database Systems
- Sr SW Engineer, Distributed Systems
- SRE, Cloud Platform
- Staff SW Engineer, Cloud Platform
- Staff SW Engineer, Database Systems

<https://zilliz.com/careers#open-positions>

# LET'S STAY CONNECTED!



<https://milvus.io/discord>



<https://github.com/milvus-io/milvus>



<https://x.com/milvusio>



<https://www.linkedin.com/company/the-milvus-project>

