



Search Medium

Write



★ Member-only story

# Vector Databases: Taking Data Revolution to The Next Level

Here's how AI-centric vector databases are gaining their place in modern software stacks.



The PyCoach · Following

Published in Geek Culture · 5 min read · May 30, 2022



247



1



The internet contains a huge amount of data in different forms. In the past, this data was mostly structured, but as the internet grew, unstructured data such as photos, audio, text, and video files became more common.

Analysts estimate that 80-90% of any organization's data is unstructured, so how can we deal with this unstructured data?

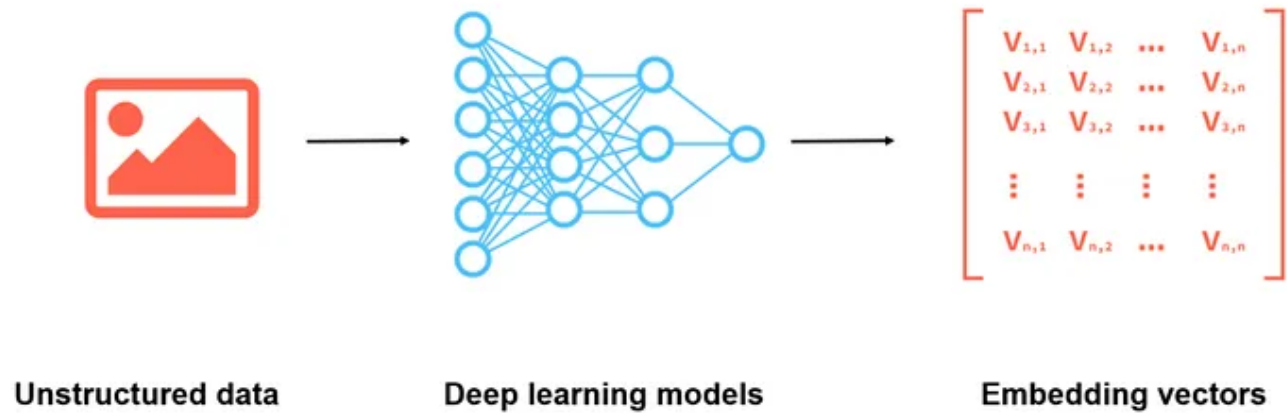
Traditional relational databases and NoSQL databases struggle to analyze unstructured data especially when it comes to doing it in real-time. Here's when vector databases can help. They were built to manage massive embeddings vectors converted from unstructured data.

In this article, we'll learn more about vector databases, how they can help us manage unstructured data, and the solutions they offer.

## **Vector Databases vs Traditional Databases**

Relational databases store items with pre-defined relationships between them. These databases typically organize data in tables with columns and rows. This is known as structured data and is generally easy to search and analyze. In contrast, unstructured data is more complex and requires more work to understand.

Machine learning and deep learning models can help us understand this unstructured data by transforming it into vector embeddings. These embeddings are high dimensional vectors that describe complex data as numerical values in different dimensions.



Source: Milvus

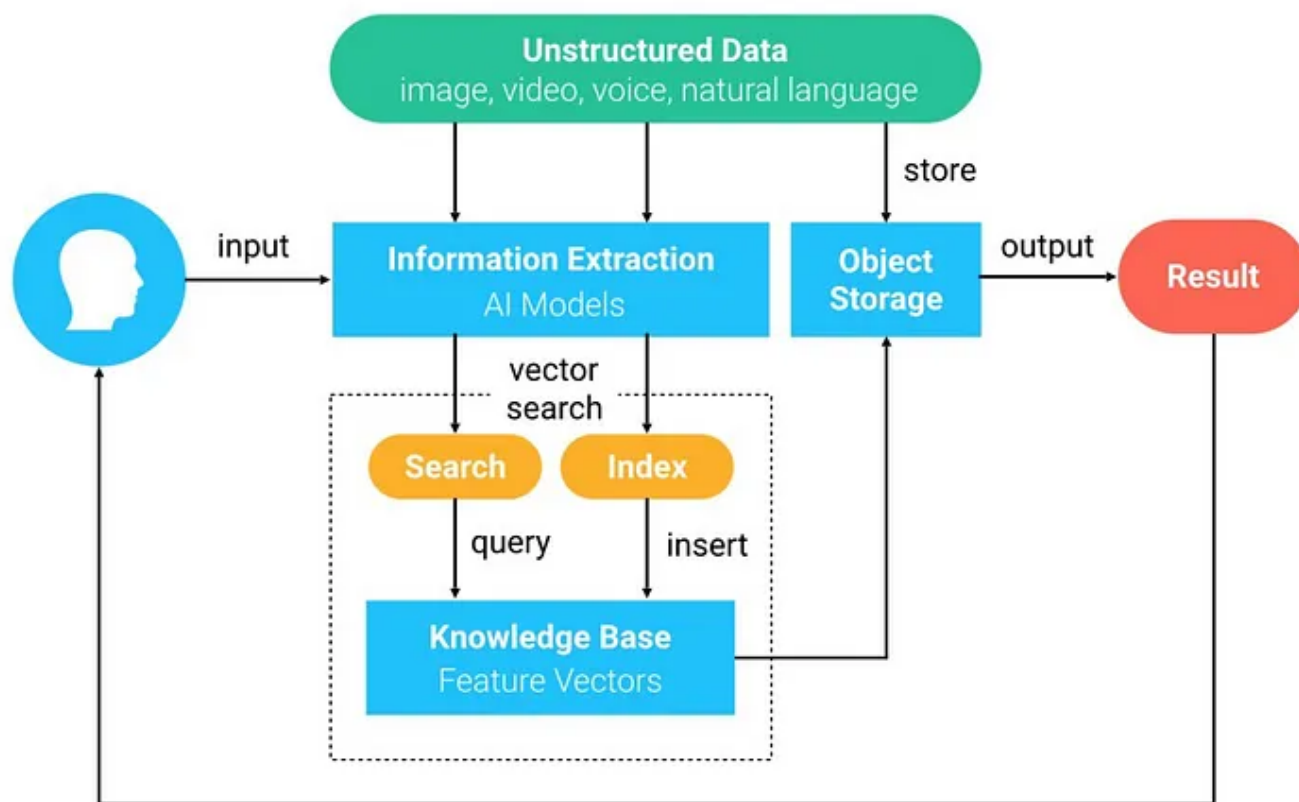
Vector databases are designed for search and data mining.

This search involves querying data as we'd do with a relational database. However, a relational database retrieves results that are an exact match, while a vector database offers more complex search capabilities.

One of the types of search that vector databases excel at is similarity search (aka vector search). Similarity search consists in finding the most similar item to the one we're looking for. This task is called nearest-neighbor search because the similar items we get are actually near-neighbor matches. This list of items wouldn't exist if we weren't performing a similarity search.

To sum it up, traditional databases return results that are an exact match, while a vector database returns near-neighbor matches. In addition to that, vector databases offer good speed, accuracy, and flexibility.

This is great for product search. Say we want to buy black Nike shoes, so we go to our favorite online store and search "black Nike shoes." If the search engine was built with similarity search functionalities, we'll get similar items, in case there aren't any black Nike shoes in the inventory.



Source: Milvus

This is just one of the applications that vector databases have. Let's see some common use cases.

## Vector Databases Use Cases

### Recommendation Engines

Recommendation systems are everywhere. They're used to suggest items similar to past videos, movies, purchases, and more.

Vector databases are great for building recommender systems. Its similarity search functionality makes vector databases a good option for suggesting relevant items to users.

We could use vector databases to build a recommender system that would suggest movies that a user might like based on the user's historical movie rating data. We could also build a system that recommends products based on past purchases a customer made.

## **Semantic Search**

Semantic search is a data searching technique that not only allows us to find keywords but seeks to understand natural language as a person would do.

How does it do that? It puts our search query into context.

We can use vector databases to index vector embeddings from NLP models in order to understand the context of the text. This provides more accurate search results.

If you remember all the things you search on Google, you'll realize that we often use natural language in our queries. Semantic search is then necessary for a search engine to provide relevant results.

## **Similarity Search**

Unstructured data such as images, video, and audio are hard to classify in relational databases.

There are some workarounds to deal with this using relational databases, but different people would use different criteria, which makes this more complicated.

In contrast, vector databases can analyze large-scale data in real-time. They're behind image search technologies such as autonomous cars that are able to recognize objects and phones that can recognize the face of a user.

They're also present in audio search. This helps our devices identify the name of a song or a user's voice.

## **A case study of a vector database**

There is an emerging market for vector databases out there. This article takes Milvus as an example and looks at its features and advantages.

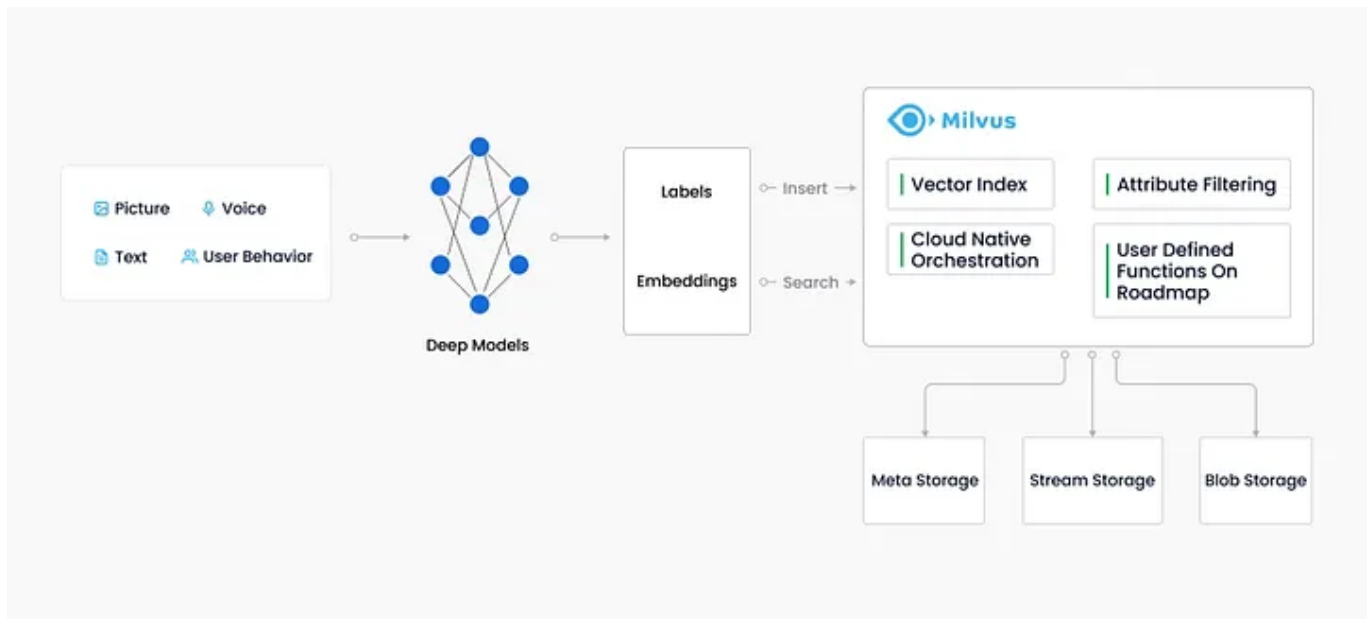
Milvus is an open-source vector database that was created with the purpose of storing, indexing, and managing embedding vectors generated by machine learning models.

Unlike relational databases that handle structured data using a pre-defined pattern, Milvus can handle embedding vectors converted from unstructured data.

Here's why Milvus is a good option:

- It offers high performance when conducting vector search on massive datasets.
- Cloud scalability and high reliability even in the event of a disruption.
- Hybrid search achieved by pairing scalar filtering with vector similarity search.

Here's the Milvus workflow.



Source: Milvus

You can check all the solutions Milvus offers [here](#).

## How to install Milvus Standalone

Milvus provides multiple ways of installation. You can easily install Milvus standalone with Docker Compose.

First, you have to download the installation file. You can download it with the following command:

```
wget https://github.com/milvus-io/milvus/releases/download/v2.0.2/milvus-standalone-docker-compose.yml -O docker-compose.yml
```

You can also download the file in the [official documentation](#).

After that, you can start Milvus by running the command below.

```
sudo docker-compose up -d
```

In case you want to check the status of the containers, run this command:

```
$ sudo docker-compose ps
```

That's it! For more information about this vector database check its [docs](#).

Learning Data Science with Python? [Get my FREE Python for Data Science Cheat Sheet by joining my email list with 10k+ people.](#)

[Data Science](#)[Python](#)[Programming](#)[Machine Learning](#)[Software Engineering](#)

---

### More from the list: "NLP"

Curated by Himanshu Birla



Jon Gi... in Towards Data ...



Jon Gi... in Towards Data ...



Jon Gi... in

**The Word2vec**



## Characteristics of Word Embeddings

★ · 11 min read · Sep 4, 2021

## The Word2vec Hyperparameters

★ · 6 min read · Sep 3, 2021

★ · 15 min read



[View list](#)



## Written by The PyCoach

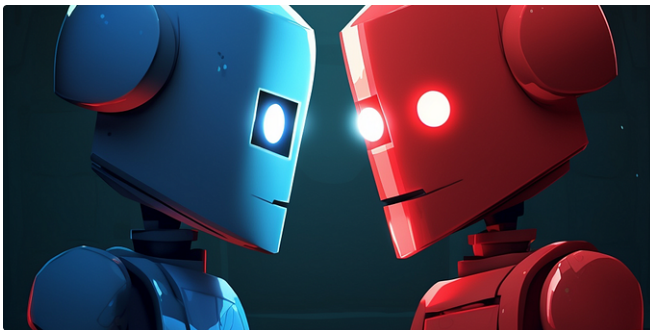
112K Followers · Writer for Geek Culture

Following



My ChatGPT Course - ChatGPT Unleashed: Master GPT-4 & Prompt Engineering: [bit.ly/chatgpt-pycoach](https://bit.ly/chatgpt-pycoach)

### More from The PyCoach and Geek Culture

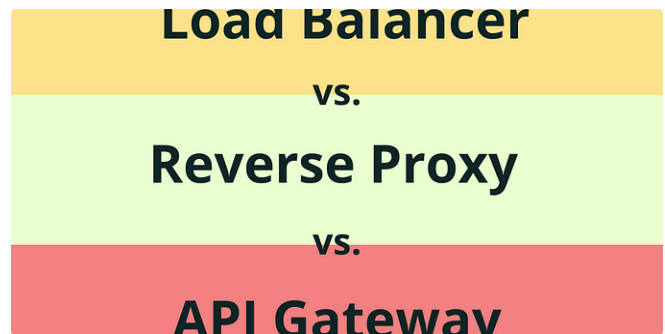


The PyCoach in Artificial Corner

## Bye-bye ChatGPT: AI Tools As Good As ChatGPT (But Few Peopl...

Go beyond ChatGPT with these powerful AI tools.

★ · 5 min read · Apr 26



Arslan Ahmad in Geek Culture

## Load Balancer vs. Reverse Proxy vs. API Gateway

Understanding the Key Components for Efficient, Secure, and Scalable Web...

12 min read · May 17



8.1K



170



1.7K



10



Farhan Tanvir in Geek Culture

## 7 Useful Java Libraries You Should Use in Your Next Project

Power up your Java development

★ · 5 min read · Aug 30



190



4



475



6



The PyCoach in Artificial Corner

## How I'd Learn AI from Scratch (If I Could Start Over)

Here's how to learn AI in 2023.

★ · 5 min read · 5 days ago



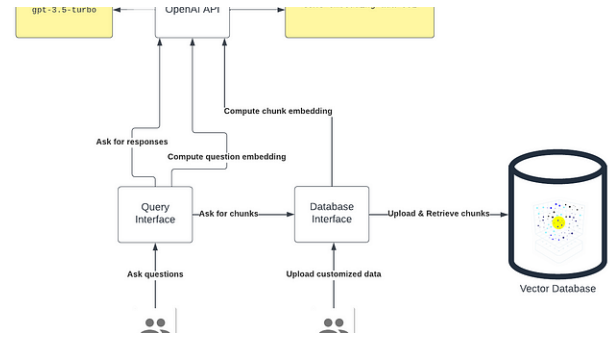
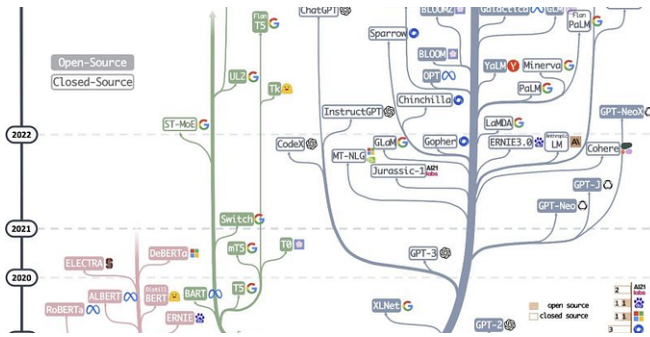
475



6

[See all from The PyCoach](#)[See all from Geek Culture](#)

## Recommended from Medium



Haifeng Li

## A Tutorial on LLM

Generative artificial intelligence (GenAI), especially ChatGPT, captures everyone's...

15 min read · Sep 14



372



...



Ruolin in Better Programming

## Enhancing ChatGPT With Infinite External Memory Using Vector...

Allowing GPT to access infinite external memory

14 min read · Apr 10



1.6K

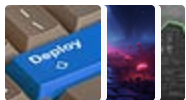


26



...

### Lists



#### Predictive Modeling w/ Python

20 stories · 452 saves



#### Practical Guides to Machine Learning

10 stories · 519 saves



#### Coding & Development

11 stories · 200 saves



#### General Coding Knowledge

20 stories · 402 saves





Han HELOIR, Ph.D. in Artificial Corner

## MongoDB and Langchain Magic: Your Beginner's Guide to Setting...

Introduction:

★ · 7 min read · Sep 12



1.4K



12



Ryan Nguyen in Towards AI

## So, You Want To Improve Your RAG Pipeline

Ways to go from prototype to production with LlamaIndex

★ · 9 min read · Sep 27



176



2



NoCode AI in AI monks.io

## Understanding Vector Database and Indexing—Visual Guide for...

What is vectors databases? Let's break it down into simple terms and explore it visually.

★ · 6 min read · May 28



33



1



ModelZ

## Do we really need a specialized vector database?

With the popularity of LLM (Large Language Model), vector databases have also become ...

11 min read · Jun 22



19

[See more recommendations](#)