Vector Database

A **Vector Database** is a type of database built to store and search **vector embeddings**..
Vectors are numerical lists that represent the meaning of data like words, sentences, images, or audio. These embeddings are created by machine learning models, like BERT or CLIP.

Instead of matching exact words or values, vector databases allow you to search by meaning. This is done using a technique called **nearest neighbor search**, which finds vectors that are mathematically close.

Common distance metrics include cosine similarity and Euclidean distance. Traditional databases (like SQL) are not optimized for this kind of search, so vector databases were built to handle it efficiently. Popular vector databases include **Pinecone**, **FAISS**, **Weaviate**, and **Milvus**.

Vector databases are commonly used in semantic search, recommendation engines, chatbots, and Retrieval-Augmented Generation (RAG) systems.
They can scale to millions of vectors and still give fast search results.

They are a key part of modern AI systems that aim to understand context, intent, and relationships. Without vector databases, semantic search and many LLM applications would be much slower and less accurate.

By enabling fast, meaningful comparisons, vector databases help make AI systems smarter and more human-like in how they retrieve information.