

Vector Databases

Instructor

Dipanjan Sarkar

Head of Community & Principal AI Scientist at Analytics Vidhya

Google Developer Expert - ML & Cloud Champion Innovator

Published Author



Outline

- Unstructured Data
- What are vectors?
- Why are vectors powerful?
- What are vector databases?
- How do vector databases store data?
- Why do vector databases index data?

Outline

- Vector database indexing techniques
- How does Vector DB search works?
- Approximate Nearest Neighbor Search
- Types of Search Mechanisms in Vector DBs
- Popular Vector Databases
- Vector Databases in LangChain

Unstructured Data

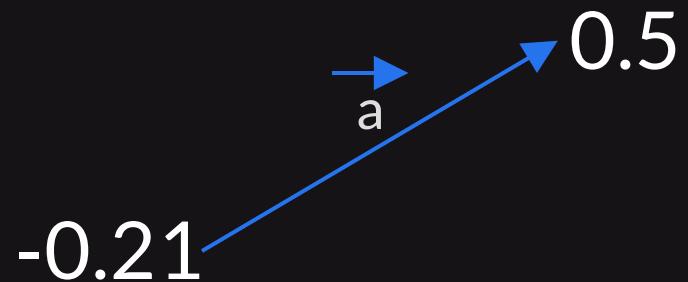
~80%

of the generated
data will be
unstructured By
2025



UNSTRUCTURED DATA

What are Vectors?



Mathematically

$[-0.21, 0.5]$

dimensions = 2

ML: List of numbers

What are Vectors?

$[-0.5, -0.5]$

dimensions = 2

$[-0.5, -0.5, 1.0]$

dimensions = 3

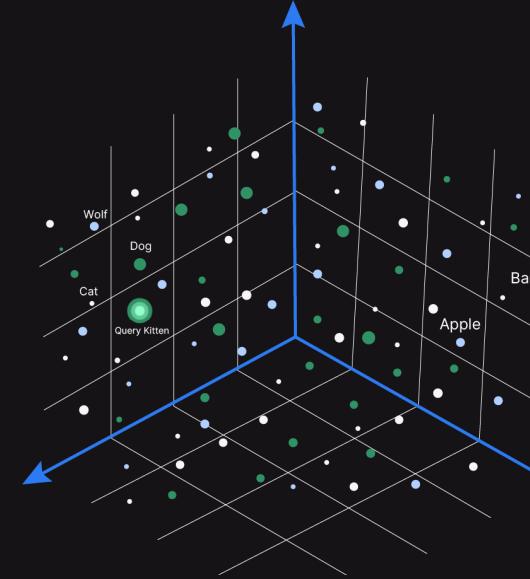
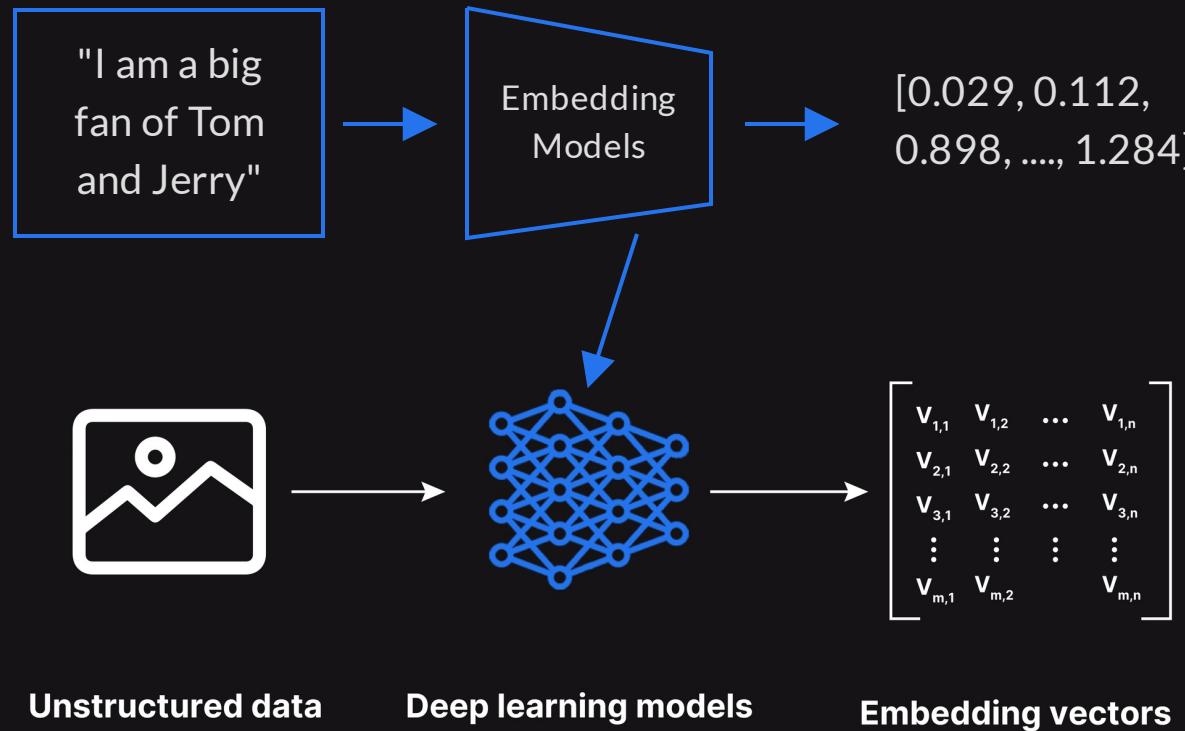
$[-0.5, -0.5, 1.0, 0.3, -0.3, 0.9, -0.9, 0.01]$

dimensions = 8

Just a list of numbers

Why are Vectors Powerful?

Vectors are a great way to represent **semantic meaning** and similar features



Distance (Apple, Banana) -> small

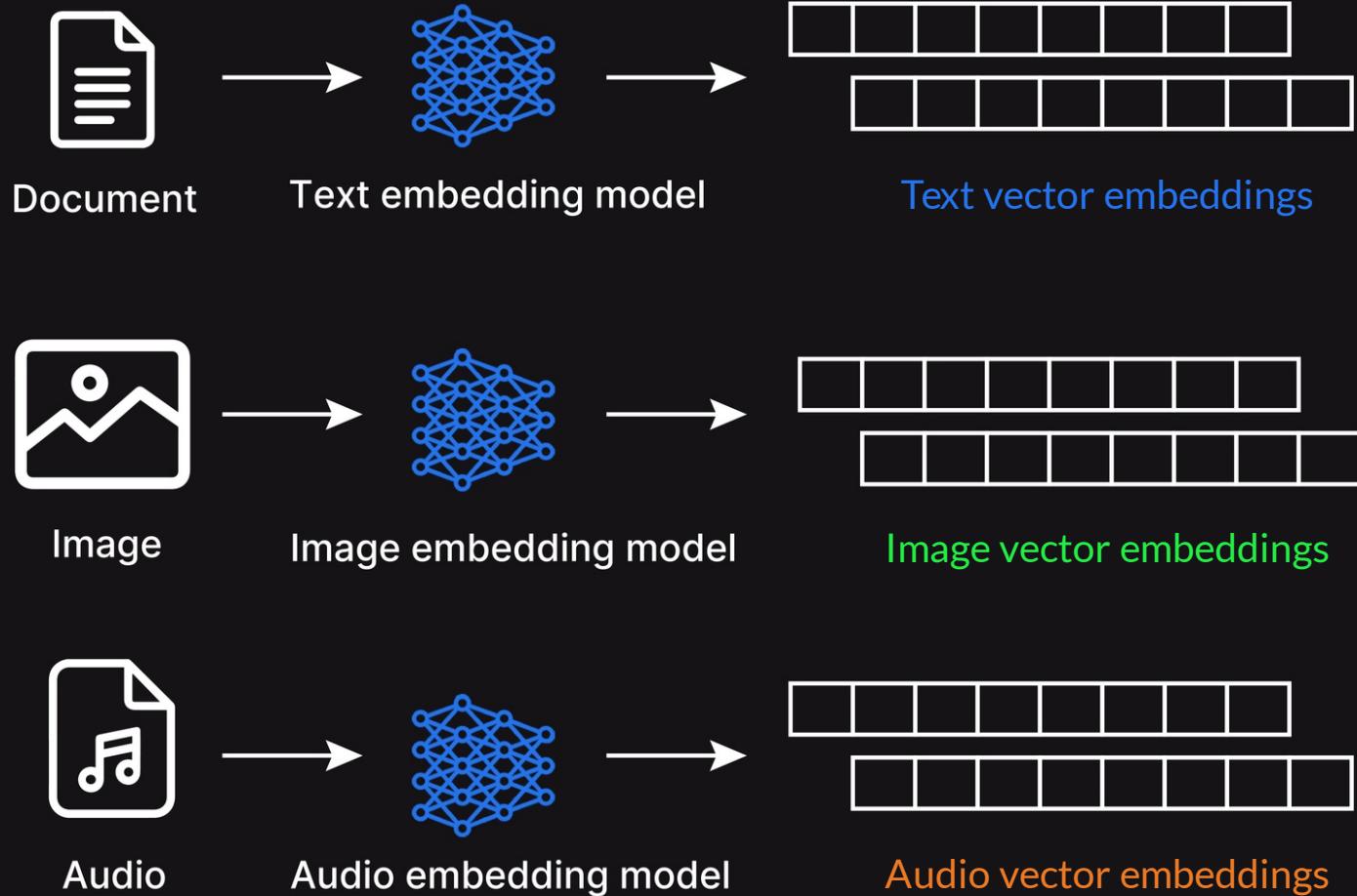
Distance (Cat, Kitten) -> small

Distance (Cat, Apple) -> large

What are Vector Databases?

- Vector databases is a DB for vector embeddings.
- It stores data as vector embeddings and enables effortless searching of high-dimensional data with ultra-low latency.
- Ideal for AI-driven applications

How do Vector Databases Store Data?



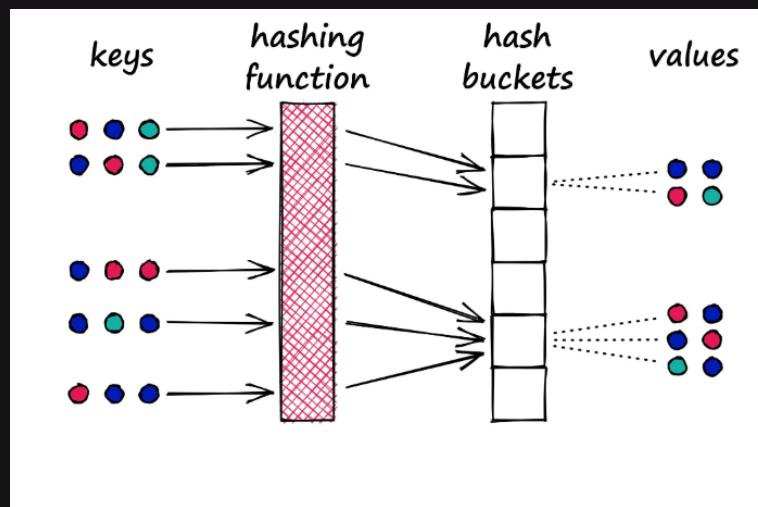
Why do Vector Databases Index Data?

- Indexing in vector databases creates a map or guide, pointing to where each vector is located on the list.
- It is a complex but essential process, enabling efficient storage, search, and retrieval of vector data.

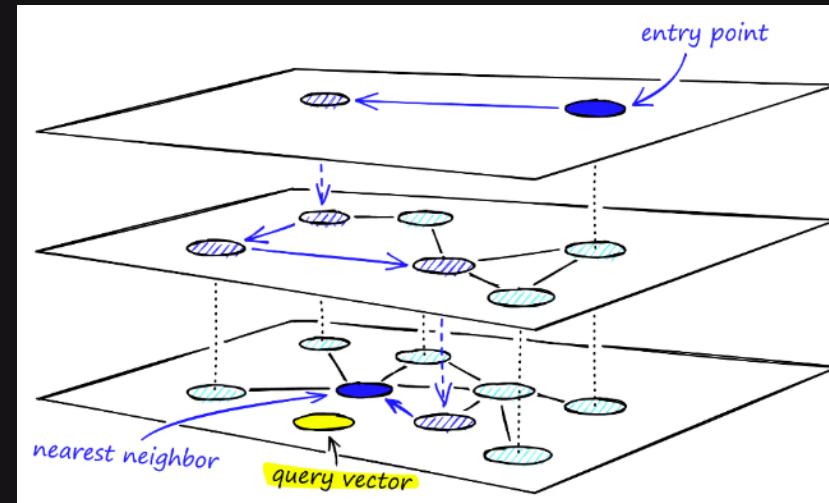
Vector Database Indexing Techniques

The vectors in a vector database are indexed for faster searches using algorithms like the ones below:

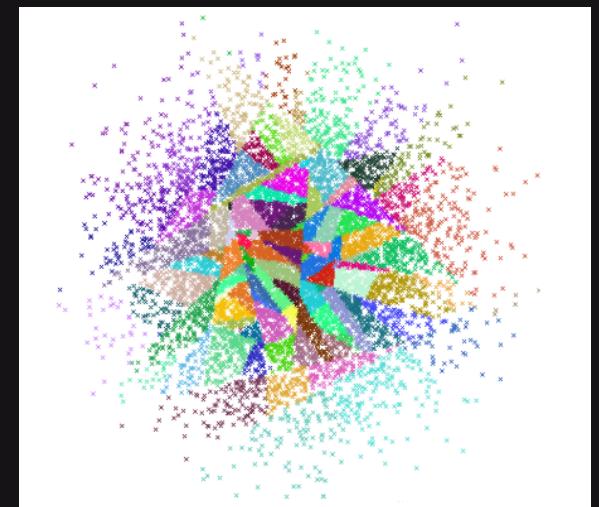
Locality-Sensitive Hashing (LSH)



Hierarchical Navigable Small World (HNSW)



Annoy (Approximate Nearest Neighbors Oh Yeah)



Vector Database Indexing Techniques



Proprietary composite index



Customized HNSW



HNSW



Flat (brute force), HNSW



HNSW + BM25 hybrid

Vector Database Indexing Techniques



Weaviate

Customized HNSW, HNSW (PQ),
DiskANN (in progress...)



Vald

NGT



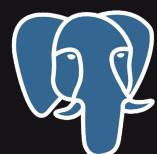
LanceDB

IVF (PQ), DiskANN (in progress...)



redis

Flat (brute force), HNSW



pgvector

IVF (Flat), IVF PQ in progress...

Vector Database Indexing Techniques



Flat, Annoy, IVF, HNSW/RHNSW (Flat/PQ),
DiskANN

How does Vector DB Search Works?

Aspect	K-Nearest Neighbors (K-NN)	Approximate Nearest Neighbors (ANN)
Algorithm	Direct computation of distances between query point and dataset points	Utilizes hashing techniques to efficiently map similar points
Accuracy	Provides exact K-nearest neighbors	Provides approximate nearest neighbors
Scalability	Computationally expensive for large datasets	Scalable to large datasets, computation cost doesn't increase linearly with dataset size
Dimensionality	Performance degrades with increasing dimensions	Efficient for high-dimensional data, performance doesn't significantly degrade

Approximate Nearest Neighbor Search - Examples

Google Search uses ANN indexing to find images similar to a query image.

Spotify recommends songs that sound similar to a track you liked.

Netflix uses ANN indexing to recommend movies and TV shows to its users.

Amazon uses hierarchical clustering to group products into different categories.

Types of Search Mechanisms in Vector DBs

Keyword Search

1. Finds relevant, useful results *when the user knows what they're looking for* and expects results that match exact phrases \ words in their search terms.
2. Does **not** require vector databases.

Semantic Search

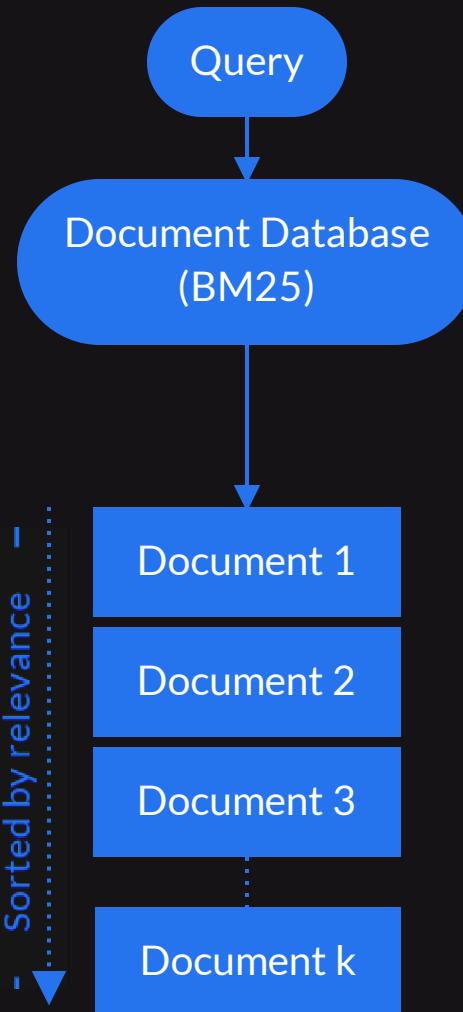
1. Finds relevant results when the *user doesn't know what exactly* they're looking for in terms of words or phrases but more in terms of meaning.
2. Requires a vector database.

Hybrid Search

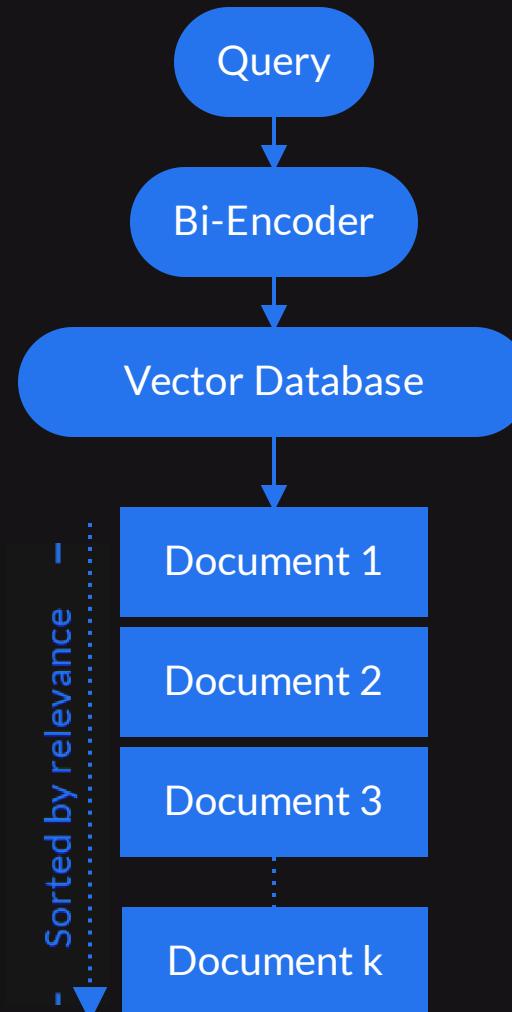
1. Typically *combines candidate results* from full-text keyword and vector searches and re-ranks them using cross-encoder models.
2. Requires both a document database and a vector database

Types of Search Mechanisms in Vector DBs

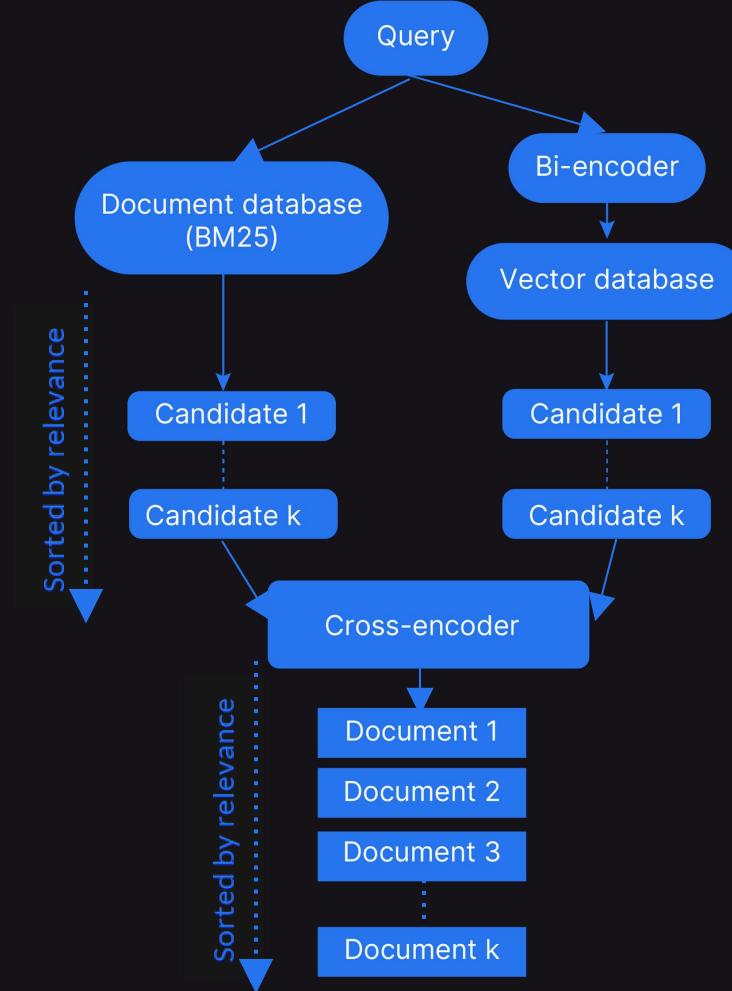
1. Keyword Search



2. Vector Search



3. Hybrid Search



Popular Vector Databases

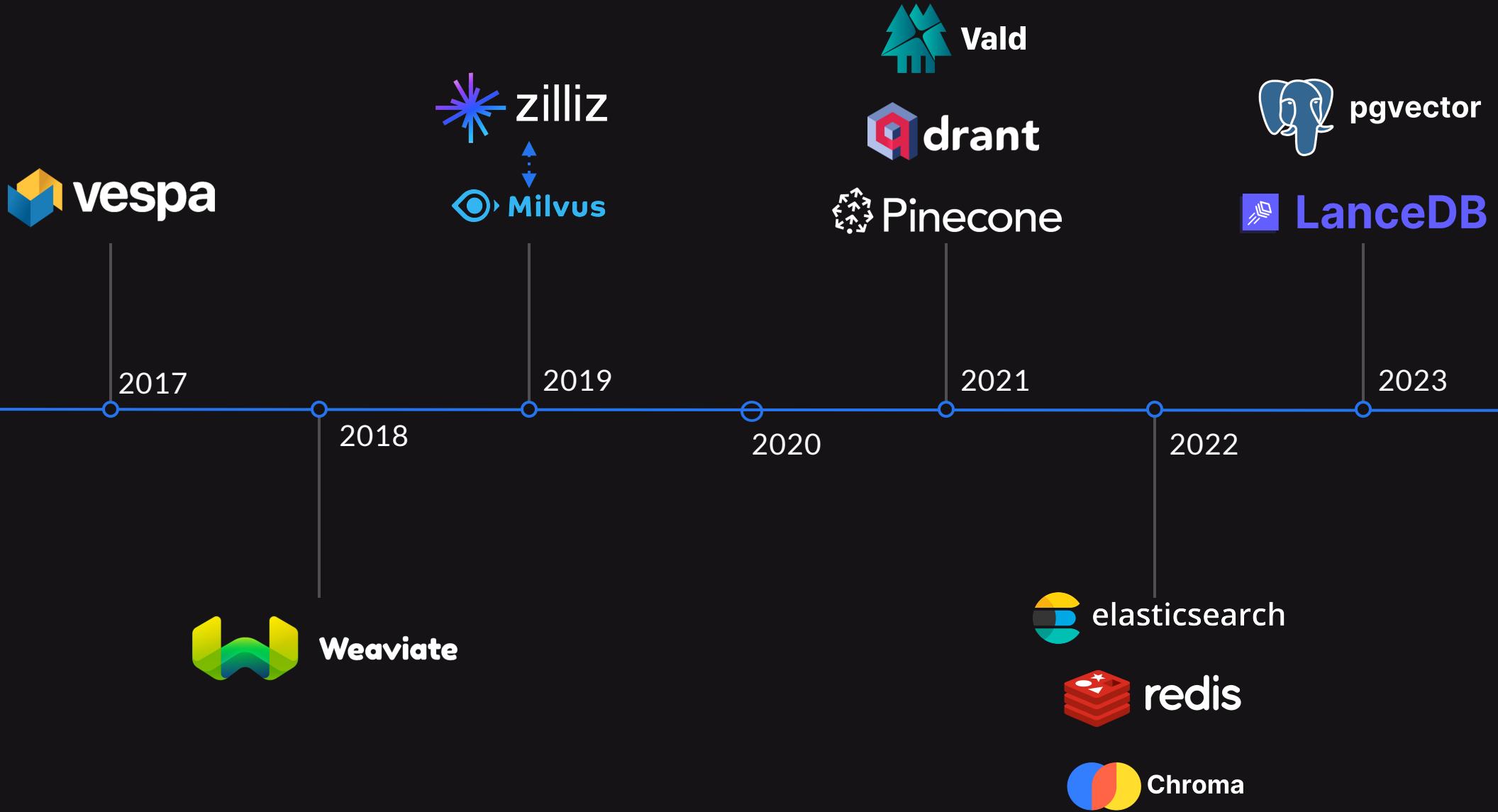
Closed-Source



Open Source/ Source Available

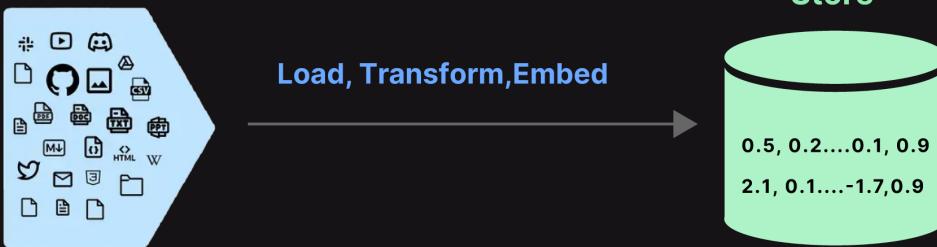


Popular Vector Databases - Timeline



Vector Databases in LangChain

1. Load Source Data



- LangChain supports API wrappers to all popular vector databases
- Integrates embedding models with vector databases to store documents and their embedding vectors
- Allows users to create, query \ read, update and delete documents and embeddings

Thank You
