CrateDB

# VECTOR DATABASES

## 5 ESSENTIAL THINGS YOU NEED TO KNOW

## 1. Definition and Importance

Vector databases are suitable for similarity search, natural language processing, and computer vision. They offer a structured approach to understanding complex patterns inside vast amounts of data.

Unlike relational databases, they use fixed-dimensional vectors to group data based on similarities, allowing for quick queries, which benefits AI-powered applications.

## 2. What should you look for when choosing a vector database?

### SIMILARITY SEARCH EFFICIENCY

This is a key function that quickly enables to find vectors in the database that are most similar to a given query vector.

### INDEXING MECHANISMS

These are key for quick retrieval of data. The database needs to offer efficient and reliable indexing mechanisms suitable for high-dimensional vector data.

### SCALABILITY

Efficiently handling large amounts of data without slowing down is crucial. The database must scale horizontally and vertically to meet the growing data demands.

### STREAMLINED DATA MANAGEMENT

Seamlessly integrating various data types like time series, geospatial, JSON, and full-text search eliminates the need for multiple systems. Plus, it offers powerful vector search capabilities.

## 3. Top Applications of Vector Databases

Potential use cases of vector storage and similarity search across several **industries and applications:**

- **E-commerce recommendations**
- **Chatbots & customer support**
- **Multimodal search**
- **Generative AI**

## 4. Combining Vector Data with Other Types of Data

To combine vector data with other types of data (structured, semi-structured, and unstructured), a **multi-model database** with a vector store suffices for most cases. It offers vital features like vector and similarity search.

Another benefit is the ability to make complex queries, joins, aggregations, and full-text searches.

## 5. CrateDB Vector Store

CrateDB is an open-source, multi-model, and distributed database that offers high performance, scalability, flexibility, and the capability to store and search vector data to accelerate AI projects.

```
SELECT text, _score
FROM word_embeddings
WHERE knn_match(embedding,[0.3, 0.6, 0.0, 0.9], 2)
ORDER BY _score DESC;
```

Statement | Result

```
|-----------------------|--------|
|         text          | _score |
|-----------------------|--------|
|Discovering galaxies   |0.917431|
|Discovering moon       |0.909090|
|Exploring the cosmos   |0.909090|
|Sending the mission    |0.270270|
|-----------------------|--------|
```

Statement | Result

Interested to know more?

**Discover CrateDB for Vector Data**