> **Document type**: Concepts guide
> **Audience**: Developers and data scientists
> **Purpose**: This document explains the core concepts, functionality, and use cases of vector databases.

Understand vector databases

Overview

Vector databases store and index data as high-dimensional vectors for semantic similarity search across unstructured data like text, images, and audio. Unlike traditional relational databases that rely on exact matching of structured data, vector databases mathematically represent meaningful features and relationships, which lets you find items that are conceptually similar rather than identical.

With vector databases, you can perform queries like "find images with similar landscapes to this mountain sunset" or "locate documents with concepts related to machine learning" without requiring exact keyword matches. For example, you can use vector databases to build recommendation systems that suggest products based on similarities customers might not explicitly search for.

How vector databases work

Vector databases operate through the following key processes:

- **Vectorization:** Vector embeddings are numerical arrays that represent the semantic meaning of data in a mathematical form. Specialized embedding models such as CLIP for images, GloVe for text, or Wav2vec for audio transform unstructured data into vector embeddings.
- **Vector representation:** Each vector consists of hundreds or thousands of numerical dimensions that correspond to specific features within the data.
- **Vector indexing:** Algorithms like Hierarchical Navigable Small World (HNSW) or Inverted File Index (IVF) organize vectors for fast retrieval.
- **Semantic similarity search:** Queries use distance metrics to locate related items based on mathematical proximity in vector space.

Benefits of vector databases

Vector databases offer several key advantages over traditional database systems:

- Discovery of semantically similar content beyond keyword matches
- Fast large-scale similarity search across billions of vectors
- Support for multiple data types like text, images, and audio
- Ability to search for concepts that traditional databases cannot express

Common use cases

Vector databases power the following AI and data applications:

- Semantic search engines
- Recommendation systems
- Image recognition
- Anomaly detection
- Retrieval Augmented Generation (RAG), where vector databases store document chunks for relevant information retrieval