

VECTOR DATABASES

Hugo Mendonça, Igor Corrêa e João Guilherme Cavalcanti

Prof. Flávio Santos

04/08/2025





O Que é um Banco de Dados Vetorial?

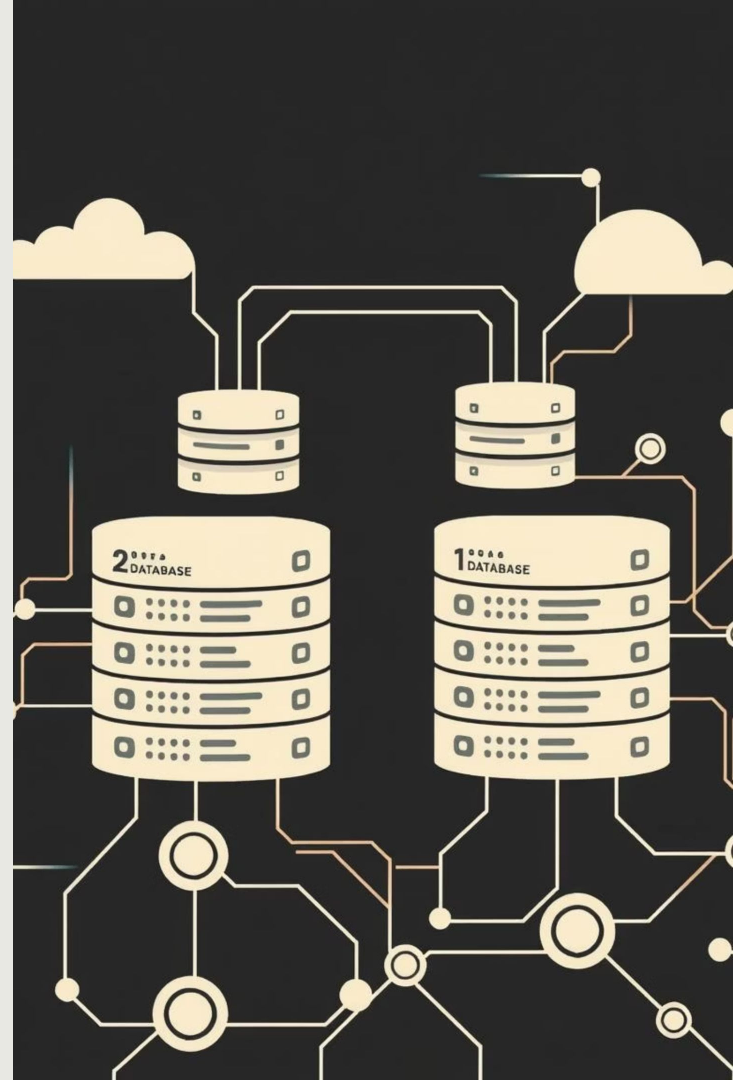
Um Banco de Dados Vetorial é um SGBD especializado em armazenar, gerenciar e consultar vetores de alta dimensão. Diferente dos bancos de dados tradicionais, ele lida com a natureza numérica e complexa dos embeddings (representação dos dados), e busca similaridade em vez de correspondências exatas.

É como uma biblioteca onde livros são organizados por "ideias" ou "temas", cada um com uma ID numérica (vetor) que descreve seu conteúdo.

Surgimento e Evolução

A necessidade de bancos de dados vetoriais surgiu com o avanço da IA e ML, especialmente para lidar com dados não estruturados (texto, imagens, áudio, vídeo). Modelos de IA transformam esses dados em representações numéricas densas chamadas embeddings, que capturam a semântica e o contexto dos dados.

Bancos de dados tradicionais não são otimizados para operações de similaridade em alta dimensão. Bancos de dados vetoriais preenchem essa lacuna, oferecendo algoritmos especializados para busca de vizinhos mais próximos (ANN) de forma eficiente, proporcionando maior escalabilidade para aplicações de IA.



Principais Características



Armazenamento de Vetores

Capacidade de armazenar vetores de alta dimensão e representações numéricas de dados não estruturados.



Indexação Eficiente

Algoritmos de indexação especializados para busca rápida de vizinhos próximos, com pequena perda de precisão.



Métricas de Similaridade

Suporte a métricas como Similaridade de Cosseno, Distância Euclidiana e Distância de Produto Interno.



Escalabilidade

Projetados para escalar horizontalmente, lidando com trilhões de vetores e milhões de consultas por segundo.



Manipulação de Metadados

Armazenamento de informações adicionais (ID, categoria, data) para contexto e filtragem de resultados.

Tipos de Dados: Embeddings

No coração de um banco de dados vetorial estão os embeddings: representações numéricas de dados em um espaço vetorial de alta dimensão. Eles capturam o significado semântico e as relações contextuais dos dados originais.

Embeddings de Texto

Palavras como "rei" e "rainha" têm vetores próximos. Frases com significados semelhantes, mesmo com palavras diferentes, também têm embeddings próximos. A frase "O gato está sentado no tapete" terá um ID próximo à frase "O felino assentou-se no carpete".

Embeddings de Imagem

Imagens de cachorros terão vetores mais próximos entre si do que imagens de carros, independentemente da raça, cor ou posição.

Tipos de Dados Suportados Além dos Vetores

Embora os vetores sejam primários, eles são quase sempre acompanhados por metadados. Metadados são dados estruturados ou semi-estruturados que fornecem informações adicionais sobre o item que o vetor representa, cruciais para contexto, filtragem e agregação.

Exemplos de Metadados

- Para texto: texto original, autor, data, categoria, tags.
- Para imagem: URL, descrição, data de captura, localização, fotógrafo.
- Para produto: nome, preço, ID, cor, tamanho, marca.

Usos dos Metadados

- Filtragem: Restringir a busca por similaridade a um subconjunto de dados (ex: "documentos de tecnologia").
- Contexto: Recuperar informações legíveis por humanos (texto original, URL da imagem).
- Organização: Agrupar ou categorizar vetores no banco de dados.

A combinação de vetores para busca por similaridade e metadados para filtragem e contexto torna os bancos de dados vetoriais poderosos e flexíveis.

Comparação com Bancos de Dados SQL

Bancos de dados vetoriais e SQL são ferramentas complementares, otimizadas para diferentes dados e padrões de acesso.

	Banco de dados SQL	Banco de dados vetorial
Modelo de Dados	Estruturado, tabelas, esquemas rígidos.	Vetores de alta dimensão e metadados.
Tipo de Dados	Numéricos, strings, datas. Foco em dados tabulares.	Vetores (embeddings) de dados não estruturados.
Tipo de Consulta	Consultas exatas, junções, agregações (SQL).	Busca por similaridade (vizinhos mais próximos, ANN).
Principal Uso	Sistemas transacionais (OLTP), relatórios.	Busca semântica, recomendação, IA/ML, RAG.

Quando Usar Cada Um?

Banco de Dados SQL

- Alta consistência e garantias transacionais (ex: sistemas bancários).
- Dados tabulares com relações bem definidas (usuários, histórico).
- Consultas exatas, junções complexas e agregações precisas.
- Prioridade máxima na integridade dos dados e precisão das buscas.

Banco de Dados Vetorial

- Encontrar itens por significado ou similaridade (imagens de cachorros).
- Grandes volumes de dados não estruturados convertidos em embeddings.
- Sistemas de recomendação, busca semântica, detecção de anomalias.
- Velocidade na busca por similaridade em escala.

Usando Ambos Juntos

Na maioria das aplicações modernas de IA, a abordagem mais eficaz é usar ambos os tipos de bancos de dados.



SQL para Dados Estruturados

Gerencia dados de usuário (perfis, permissões, histórico de conta).



Vetorial para Funcionalidades de IA

Armazena embeddings e metadados para busca semântica e recomendações.

Ex: Um e-commerce usa SQL para pedidos e banco de dados vetorial para recomendar produtos.

Casos de Uso de Bancos de Dados Vetoriais



Busca Semântica

Encontra resultados baseados no significado da consulta, não apenas palavras-chave exatas (ex: "bebidas quentes para o inverno" retorna café, chá).



Detecção de Anomalias

Identifica padrões incomuns ou "outliers" em grandes conjuntos de dados (ex: fraude financeira, intrusões de rede).



Busca de Imagens e Vídeos

Pesquisar conteúdo visual por similaridade ou descrições de texto (ex: "vídeos de pessoas andando de bicicleta na praia").

OBRIGADO!