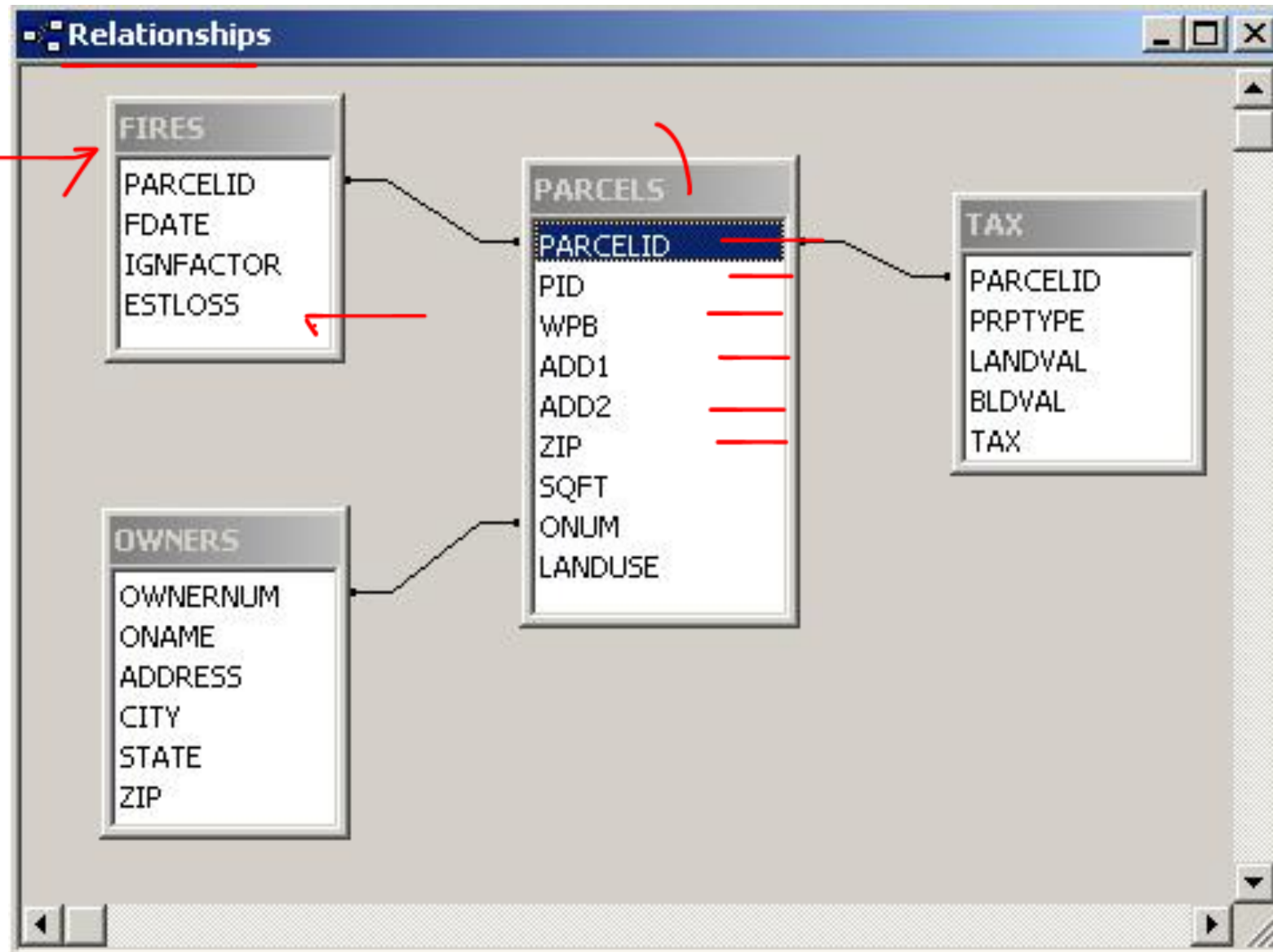


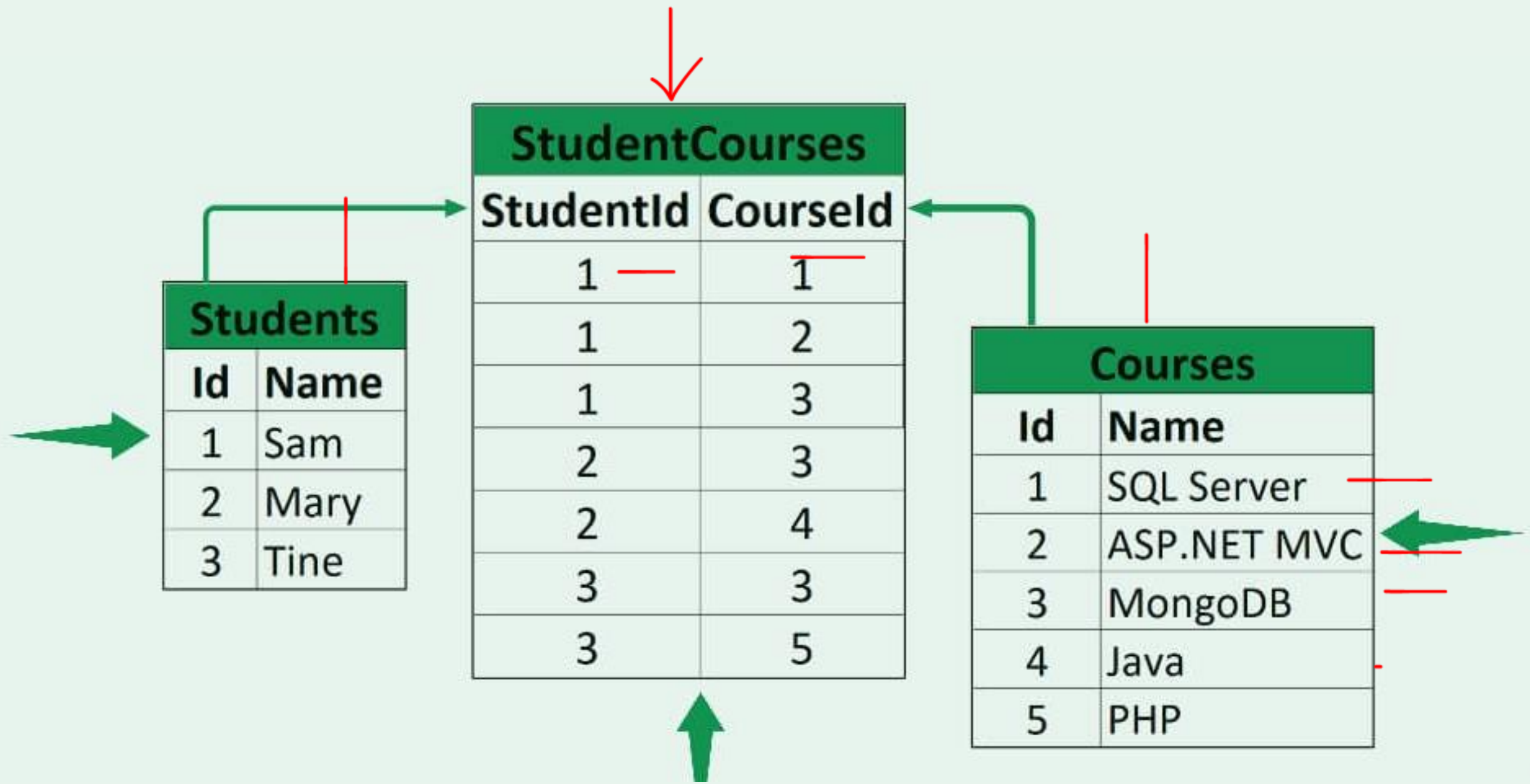


# Vector DataBase

A simple introduction to newbie

- Nội dung slide đã được đơn giản hóa hết mức có thể nhằm phù hợp với Mì AI.
- Một số phần có thể thiếu thông tin, mong bạn thông cảm!





# BLOB

---

**B**INARY  
**L**ARGE  
**O**BJECT



✕ ✓ fx

= Sql.Database(".", "AdventureWorks2012", [Query="SELEC

	ProductID	Name	ThumbNailPhoto	LargePhoto
1	771	Mountain-100 Silver, 38	Binary	Binary
2	772	Mountain-100 Silver, 42	Binary	Binary
3	773	Mountain-100 Silver, 44	Binary	Binary
4	774	Mountain-100 Silver, 48	Binary	Binary
5	775	Mountain-100 Black, 38	Binary	Binary
6	776	Mountain-100 Black, 42	Binary	Binary



# Vector database

(noun)

indexes, stores, and provides access to  
structured or unstructured data —  
alongside its vector embeddings.

It allows users to find and retrieve similar  
objects quickly at scale in production.

## Vector Database Là Gì?

Hãy tưởng tượng một thư viện khổng lồ không sắp xếp sách theo tên hay tác giả, mà theo nội dung và ý nghĩa. Vector Database làm điều tương tự với dữ liệu. Nó là một loại cơ sở dữ liệu chuyên dụng để lưu trữ, quản lý và tìm kiếm dữ liệu dưới dạng các "vector embeddings" - những chuỗi số đại diện cho ý nghĩa ngữ nghĩa của dữ liệu gốc.



### Văn bản

"Mèo là loài động vật đáng yêu."



### Hình ảnh

Một bức ảnh chú mèo Ba Tư.



### Âm thanh

Tiếng mèo kêu "meow".



## Chuyển đổi thành Vector Embeddings

[0.12, 0.98, -0.34, ..., 0.67]

Mỗi vector là một điểm trong không gian đa chiều, nơi khoảng cách giữa các điểm thể hiện sự tương đồng về ngữ nghĩa.

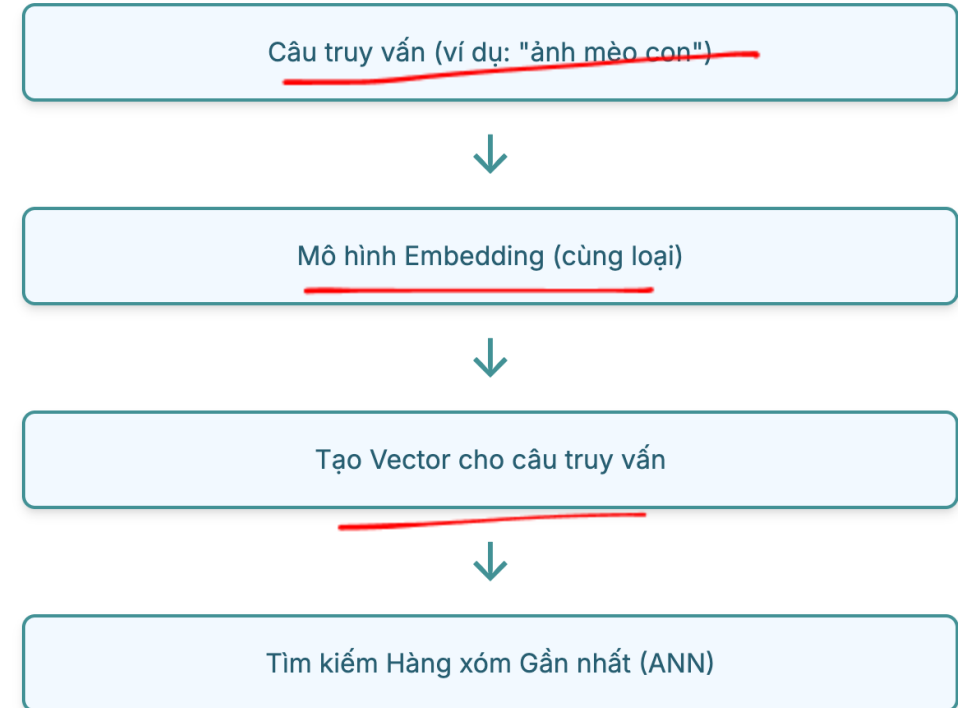


## Cách Thức Hoạt Động

### 1. Lập Chỉ Mục (Indexing)

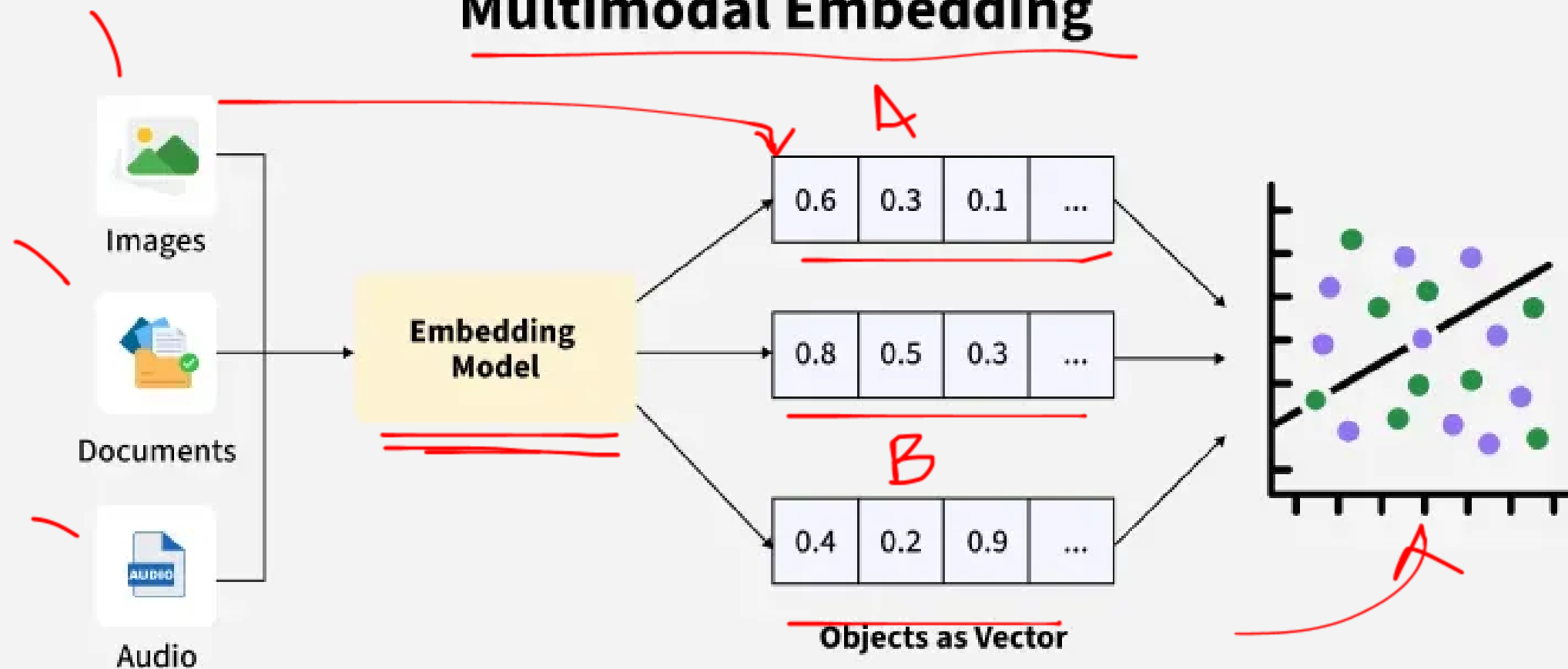


### 2. Truy Vấn (Querying)



Cốt lõi của việc truy vấn là tìm kiếm sự tương đồng. Vector DB sử dụng các thuật toán như ANN (Approximate Nearest Neighbor) để nhanh chóng tìm ra các vector trong cơ sở dữ liệu gần với vector truy vấn nhất, thay vì tìm kiếm đối sánh chính xác.

# Multimodal Embedding



## So Sánh: Vector DB và Relational DB

Dù cùng là cơ sở dữ liệu, chúng được thiết kế cho những mục đích hoàn toàn khác nhau. Dưới đây là sự khác biệt cơ bản về cấu trúc, loại dữ liệu và cách chúng truy vấn thông tin.

### Vector Database

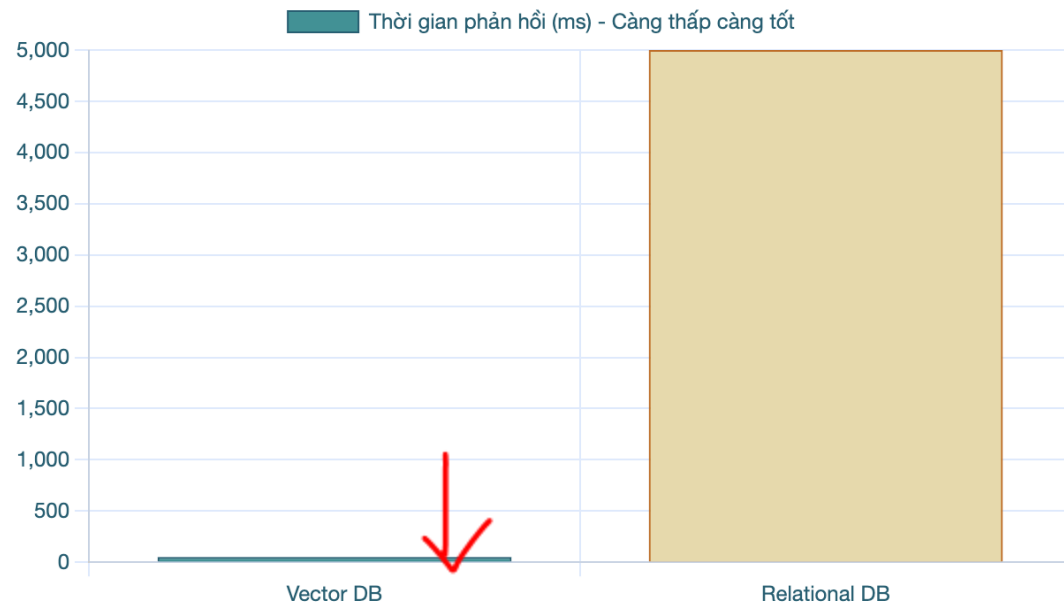
- **Cấu trúc:** Lưu trữ vector đa chiều.
- **Dữ liệu:** Tối ưu cho dữ liệu phi cấu trúc (text, image, audio).
- **Truy vấn:** Dựa trên sự tương đồng ngữ nghĩa (similarity search).
- **Câu hỏi ví dụ:** "Tìm các tài liệu có nội dung giống với đoạn văn này."
- **Logic:** "Gần giống", "Liên quan đến".

### Relational Database (ví dụ: SQL)

- **Cấu trúc:** Bảng (rows và columns) có cấu trúc chặt chẽ.
- **Dữ liệu:** Tối ưu cho dữ liệu có cấu trúc (số, chuỗi, ngày tháng).
- **Truy vấn:** Dựa trên đối sánh chính xác (exact match).
- **Câu hỏi ví dụ:** "Lấy tất cả người dùng có tên là 'An' và tuổi > 25."
- **Logic:** "Bằng", "Lớn hơn", "Nhỏ hơn".

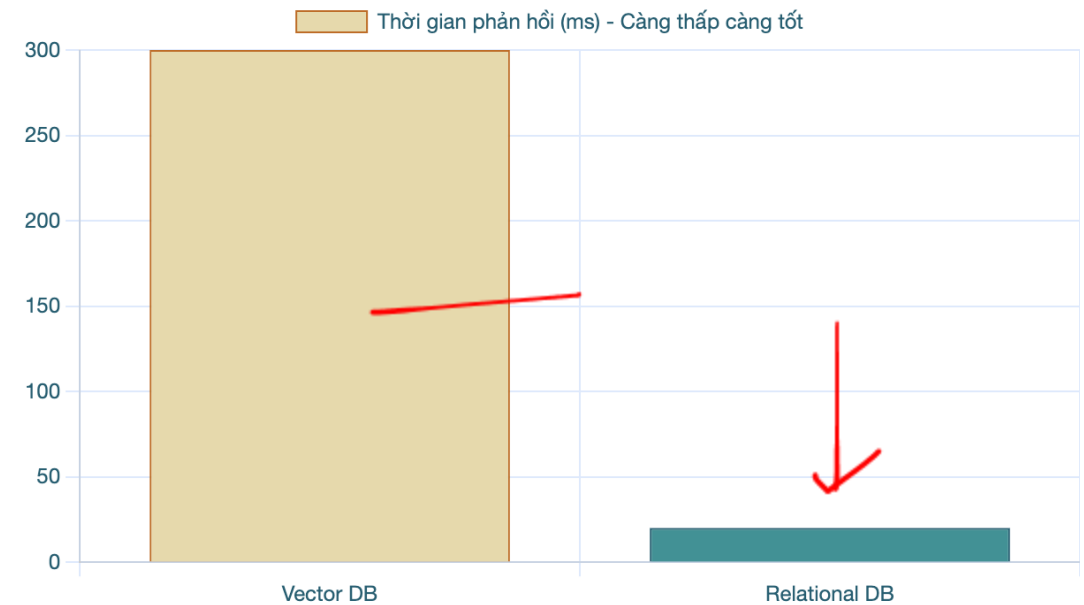
## Hiệu Suất Tìm Kiếm Tương Đồng

Với các tác vụ tìm kiếm dựa trên ý nghĩa, Vector DB vượt trội hơn hẳn.



## Hiệu Suất Truy Vấn Chính Xác

Relational DB được tối ưu cho việc lọc dữ liệu có cấu trúc với điều kiện chính xác.



# Các Ứng Dụng Thực Tế

Sự trỗi dậy của AI và học máy đã đưa Vector Database trở thành một công cụ không thể thiếu trong nhiều lĩnh vực, giúp các ứng dụng "hiểu" được dữ liệu theo cách gần giống con người.



## Tìm kiếm Hình ảnh/Video

Tìm kiếm hình ảnh bằng một hình ảnh khác hoặc mô tả văn bản ("tìm ảnh bãi biển hoàng hôn").



## Hệ thống Gợi ý

Gợi ý sản phẩm, bài hát, hoặc bộ phim dựa trên những gì bạn đã thích trước đó.



## Xử lý Ngôn ngữ Tự nhiên (NLP)

Xây dựng chatbot thông minh, phân loại văn bản, và tìm kiếm ngữ nghĩa trong tài liệu lớn.



## Phát hiện Bất thường

Phát hiện gian lận tài chính hoặc các mối đe dọa an ninh mạng bằng cách xác định các mẫu khác biệt.



## Tăng cường cho Mô hình Ngôn ngữ lớn (LLM)

Cung cấp bộ nhớ dài hạn và kiến thức chuyên ngành cho các mô hình như GPT, tránh tạo ra thông tin sai lệch.



## Nhận dạng Âm thanh

Tìm kiếm một bài hát chỉ bằng cách ngân nga một giai điệu (ví dụ: Shazam).

Open source  
(Apache 2.0 or MIT license)

## Dedicated vector databases



Source available  
or commercial



## Databases that support vector search



# Demo speed of Vector DB

