Lecture 24: PG Vector Store Intro

Why Not a Simple Vector Store?

- **Simple Vector Store** is good for **learning and experiments but** not ideal for production.
- Real projects need a **persistent and reliable vector database**.
- That's where **PGVector** comes in.

What is PGVector?

- **PGVector** is an **open-source PostgreSQL extension** for handling vector embeddings.
- It allows storing embeddings in Postgres and performing **similarity searches**.
- Built directly on PostgreSQL, making it reliable and easy to integrate if you already use Postgres

Installation & Setup

- PostgreSQL does **not include PGVector by default**.
- You need to install/enable it separately.
- Options to set it up:
  - **Docker Image/Docker Compose is the easiest** approach.
  - Local installation (platform-specific steps for Windows, Linux, and Mac).

Spring AI + PGVector:

- In Spring AI, PGVector works as a **Vector Store implementation**.
- You will need:
  - **Dependencies** → add the pgvector dependency along with Spring AI.
  - **Database Configuration** → URL, username, password.
  - **Vector Store Properties** → specify:
    - index type
    - distance type (cosine, L2, etc.)
    - dimensions
    - max document batch size

Key Takeaways:

- PGVector lets you **store and query embeddings inside PostgreSQL**.
- Best to use in production where persistence and scalability matter.

TELUSKO