Vector Databases (VectorDB)

Overview
A Vector Database (VectorDB) is a specialized data store designed to manage, index, and search high-dimensional vector embeddings efficiently. These embeddings are numerical representations of data such as text, images, audio, or code, commonly generated by machine learning models.

Why Vector Databases Are Needed
Traditional databases are optimized for exact matches and structured queries, but they perform poorly when handling semantic similarity. Vector databases enable fast similarity search, allowing applications to find conceptually related data rather than exact keyword matches.

How Vector Databases Work
1. Raw data is converted into embeddings using an embedding model.
2. Embeddings are stored along with metadata.
3. Indexing techniques (e.g., HNSW, IVF) optimize similarity search.
4. A query is embedded and compared using distance metrics like cosine similarity or Euclidean distance.

Key Features
- High-dimensional vector storage
- Approximate Nearest Neighbor (ANN) search
- Metadata filtering
- Horizontal scalability and low-latency queries

Common Vector Database Technologies
Popular vector databases include Pinecone, Weaviate, Milvus, FAISS, Chroma, and vector-enabled databases such as PostgreSQL (pgvector) and OpenSearch.

Use Cases
Vector databases power semantic search, recommendation systems, Retrieval-Augmented Generation (RAG), document similarity, anomaly detection, and multimodal AI applications.