

# Building AI-Powered Search Applications on VMware Greenplum Data Warehouse with pgvector



In the past year, embeddings have gained significant traction as a powerful technique in the fields of Machine Learning and Generative-AI. These embeddings serve as vector representations of data points, capturing their essential characteristics and features.

However, as the demand for efficient semantic search at a large scale increases, there arises a need for a robust platform capable of storing embeddings and enabling seamless search capabilities. Greenplum, a cutting-edge data warehousing solution now equipped with the performance-oriented pgvector. As an open-source extension for PostgreSQL, pgvector empowers customers to store ML embeddings, construct AI applications, and execute high-performing similarity searches.

In this talk, we'll learn how to leverage its powerful vector similarity search capabilities within Greenplum and harness its potential in combination with OpenAI models and finally discover how this integration can revolutionize the development of Image Search applications and domain-specific Chatbots.

## Key Points:

- Introduction to pgvector: the Open-source vector similarity search for Postgres.
- Store and query ML Embeddings inside Greenplum using pgvector extension.
- Perform efficient semantic similarity search at scale on Image & Text Embeddings.
- Simply set up, operate, and scale your ML-enabled applications.
- Demo 1: Build an AI-powered Chatbot for your product documentation in Greenplum combining pgvector and OpenAI API.
- Demo 2: Text-to-Image and Image-to-Image Search Using CLIP model and pgvector.

## Takeaways:

- Leverage Greenplum data warehouse as a Vector Database for large scale AI-Applications
- Store unstructured data inside a relational database as Embeddings/Vectors alongside metadata
- Process Texts & Images and build ML-applications using Greenplum's Python library and pgvector extension