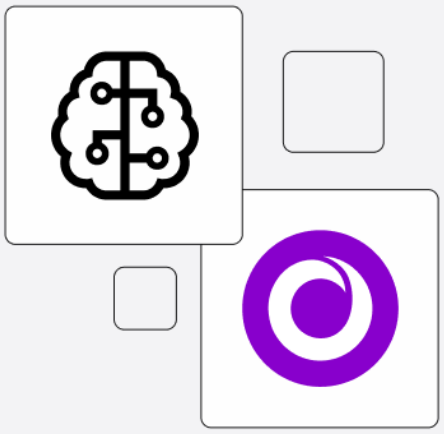
Solution Brief

Generative AI (GenAI) with SingleStoreDB

A contextual database for generative AI applications



Vector functions and so much more…

SingleStoreDB is a high-performance, scalable, SQL database and cloud service that supports **multiple data models** including JSON, time-series, full text, spatial, key-value and vector data. It can power high-performance transactional and analytical workloads together with vector capabilities in the same engine, without data movement. With its **fast data ingestion** from multiple sources and low-latency queries across all data, SingleStoreDB offers real-time capabilities few others can match.

- **Vector functions, battle tested**. SingleStoreDB has included **vector capabilities since 2017** and several customers use it for storing, processing and performing nearest-neighbor searches to power gen AI applications. SingleStoreDB supports **multiple data formats** natively (including JSON, full text search, time-series, geospatial and vectors), making it possible to power AI apps with all kinds of structured and unstructured data "under one roof".
- **Easy to get started.** SingleStoreDB offers integrations or plugins for **popular tools like OpenAI, Hugging Face, LangChain and LlamaIndex** making it easy to get started quickly. SingleStoreDB also offers **hybrid search capabilities** that can combine semantic and full-text search.

Real-time AI with full context

Use all data relevant to your company. Combine vector embeddings from text, images, audio, video, etc., with other kinds of data including logs, stock market data, clickstream or sensor data. All kinds of structured and unstructured data can be co-located and queried using SingleStoreDB — including vectors, JSON, time-series, text, SQL and geospatial data.

Fast data ingestion from other sources. SingleStoreDB supports a wide range of data sources and connectors, allowing users to ingest data from diverse systems including other databases, HDFS, message queues, log files, cloud storage (Amazon S3) and streaming data platforms like Confluent Kafka.

Rich query language and powerful analytics

Powerful SQL. Metadata filtering, joins, aggregates, subqueries, window functions and other language features. Ability to **rerank semantic search results** based on real-time data.

Hardware acceleration. Built-in parallelization and Intel Single Instruction Multiple Data (SIMD)-based vector processing for fast, reliable vector similarity matching.

High performance, even for complex queries. Low latency queries (~milliseconds) matching top data warehouses on performance benchmarks such as TPC-H and TPC-DS.

A simpler and more effective vector database

Simple. Eliminate the complexity, licensing costs or extra training requirements of a pure vector database.

Fast K-Nearest-Neighbor search. Use 'order by/limit k' queries using 'dot_product' and 'euclidean_distance' metrics, combined with arbitrary SQL for metadata filtering. **Re-ranking semantic search results** is easy with support for 'dot_product' and 'match'.

Notebooks **(Preview)** Quickly prototype and deploy with **Notebooks** that combine the power of SQL and Python.

Gen AI ecosystem. Ability to use platforms, plugins and libraries like **ChatGPT, LangChain, Hugging Face, llama 2**, etc. to build gen AI applications right next to where the data resides.

Enterprise ready. Get the data security, compliance and high availability appropriate for enterprise applications

Other key capabilities

Universal storage
Both rowstore and columnstore in one database.

Unlimited storage
Separation of storage and compute.

SingleStore Kai™
Up to 1,000x faster analytics on MongoDB® apps.

High availability
Your applications should stay online & be highly available.

Run anywhere
Hybrid, multi-cloud, SaaS, on-premises, Kubernetes operator.

SIEMENS THORN LUMIX.AI directly apply nyris

Learn more about SingleStoreDB and its vector capabilities.