



Komunikat prasowy

KIOXIA stworzyła bazę danych umożliwiającą wyszukiwanie wektorów wielowymiarowych o rozmiarze 4,8 miliarda elementów na jednym serwerze, przy 7,8-krotnym przyspieszeniu tworzenia indeksu dzięki GPU

Wykorzystanie biblioteki NVIDIA cuVS z technologią KIOXIA AiSAQ™ do indeksowania wektorów o wymiarach 1024 przy minimalnym użyciu pamięci DRAM

Niemcy, Düsseldorf, 16 marca 2026 – [KIOXIA Europe GmbH](#) ogłosiła dziś udaną demonstrację KIOXIA Corporation, w której osiągnięto skalowanie wyszukiwania wektorów wysokowymiarowych do 4,8 miliarda wektorów na jednym serwerze przy użyciu otwartoźródłowej technologii KIOXIA AiSAQ™ do przybliżonego wyszukiwania najbliższych sąsiadów (ANNS). Dodatkowo KIOXIA wykazała znaczące skrócenie czasu budowy indeksu dzięki przyspieszeniu GPU dzięki [NVIDIA cuVS](#). Te dwa osiągnięcia stanowią istotny postęp w zakresie rozwiązań generowania treści wspomaganego wyszukiwaniem (RAG). Trwają prace nad dalszym rozwojem rozwiązania, aby umożliwić wdrożenia na większą skalę, obejmujące ponad 4,8 miliarda wektorów.

Czas budowy indeksu na masowej bazie wektorowej to kluczowy problem dla branży. We współpracy z NVIDIA, KIOXIA zademonstrowała do 20-krotnej poprawy czasu budowy indeksu KIOXIA AiSAQ™ dla wektorów wysokowymiarowych o 1024 wymiarach oraz do 7,8-krotnej poprawy całkowitego czasu budowy indeksu (end-to-end). Ta 20-krotna poprawa oznacza redukcję z 28,4 dnia użytkowania CPU do 1,4 dnia z użyciem czterech kart [GPU NVIDIA Hopper](#) do budowania indeksu oraz skrócenie czasu z 31 dni do 4 dni w testach end-to-end.¹



Aplikacje AI mogą teraz polegać na większych wolumenach zwektoryzowanych danych sięgających dziesiątek miliardów wektorów i więcej, przechowywanych na dyskach SSD, podczas gdy sama pamięć DRAM staje się niepraktyczna nawet przy skali miliardowej. KIOXIA umożliwia wysoce skalowalną architekturę pamięci masowej dzięki technologii KIOXIA AiSAQ™, osiągając wyszukiwanie na skalę miliardową, przekraczając wymagania dotyczące opóźnień aplikacji RAG przy użyciu pojedynczego serwera zapytań w środowisku Milvus vectorDB, wspieranego przyspieszeniem GPU w procesie budowy indeksów, co sprawia, że wdrożenia na dużą skalę stają się praktyczne.

„Bazy danych wektorowych stanowią fundament dla aplikacji, które muszą rozumieć intencje, kontekst i podobieństwo w ogromnych, nieustrukturyzowanych zbiorach danych w czasie rzeczywistym” – powiedział Jason Hardy, wiceprezydent, Storage Technologies w NVIDIA. „Dzięki wykorzystaniu przyspieszonej GPU indeksacji z biblioteką NVIDIA cuVS, KIOXIA wspiera bazy danych wektorowych wysokowymiarowych, które mogą się skalować i budować indeksy z niespotykaną dotąd wydajnością.”

[Ogłoszona po raz pierwszy ogłoszony w zeszłym roku](#) technologia oprogramowania open-source KIOXIA AiSAQ™ rozwiązuje wyzwania skalowalności RAG, umożliwiając wyszukiwanie wektorowe bezpośrednio z SSD, przy zmniejszonym użyciu DRAM. Technologia KIOXIA AiSAQ™ zapewnia wysoką skalowalność, dzięki czemu jest dobrze dopasowana zarówno do środowisk wielodostępnych, jak i do wdrożeń indeksów monolitycznych na dużą skalę. Technologia wykorzystuje innowacyjny algorytm Global Index, który łączy klasteryzację hybrydową z wyszukiwaniem w grafie, zapewniając wydajne przeszukiwanie wektorów w ekstremalnej skali. Dzięki elastycznym opcjom dostrajania umożliwiającym równoważenie wydajności i skalowalności wektorów o dużej objętości, oprogramowanie KIOXIA AiSAQ™ sprawia, że wdrożenia na dużą skalę stają się bardziej dostępne i łatwiejsze do rozbudowy.

„Osiągnięcie skalowalnego wyszukiwania wektorów na skalę wielomiliardową na jednym serwerze stanowi ważny kamień milowy dla branży” – powiedział Axel Stoermann, Chief Technology Officer i Vice President w KIOXIA Europe GmbH.



„Dzięki wyszukiwaniu wektorów opartemu na dyskach SSD w technologii KIOXIA AiSAQ™ oraz przyspieszonej GPU indeksacji z użyciem NVIDIA cuVS, skracamy czas budowy indeksów z tygodni do dni – to kluczowy krok w kierunku uruchomienia systemów RAG na dużą skalę.”

KIOXIA pozostaje zaangażowana w rozwój rozwiązań AI opartych na pamięci masowej, wspierających inteligentne przetwarzanie danych na dużą skalę, i kontynuuje rozwój technologii KIOXIA AiSAQ™ w kierunku wdrożeń na skalę biliona wektorów.

Link do pobrania oprogramowania open-source KIOXIA AiSAQ™: <https://github.com/kioxia-jp/aisaq-diskann>.

###

Uwagi:

1: Łącznie przetworzono 19,66 TB danych wektorowych dla tego benchmarku. Wydajność lub wyniki testów porównawczych mogą się różnić w zależności od urządzenia hosta, warunków odczytu i zapisu, rozmiaru danych oraz innych czynników.

KIOXIA AiSAQ jest znakiem towarowym firmy KIOXIA.

Nazwy firm, produktów oraz usług mogą stanowić znaki towarowe podmiotów trzecich.

O KIOXIA Europe GmbH

KIOXIA Europe GmbH jest europejską spółką zależną KIOXIA Corporation, wiodącego na świecie dostawcy pamięci flash i dysków stałych (SSD). Od czasu wynalezienia pamięci flash NAND po dzisiejsze cieszące się uznaniem technologie pamięci flash 3D BiCS FLASH™, KIOXIA kontynuuje pionierskie rozwiązania i usługi w zakresie pamięci, które wzbogacają życie ludzi i poszerzają horyzonty społeczeństwa. Innowacyjna technologia firmy w obszarze pamięci flash 3D, BiCS FLASH™, kształtuje przyszłość przechowywania danych w zastosowaniach o dużej gęstości zapisu, w tym w zaawansowanych smartfonach, komputerach PC, systemach samochodowych, centrach danych oraz w systemach generatywnej sztucznej inteligencji.

Odwiedź naszą [stronę internetową KIOXIA](#)



Dane kontaktowe do publikacji:

KIOXIA Europe GmbH, Hansaallee 183, 40549 Düsseldorf, Niemcy

Tel.: +49 (0)211 368 77-0

E-mail: KIE-support@eu.kioxia.com

Dane kontaktowe w przypadku pytań dot. publikacji:

Lena Hoffmann, KIOXIA Europe GmbH

Tel.: +49 (0) 211 36877 382

E-mail: lena.hoffmann@eu.kioxia.com

Wydawca:

Birgit Schöninger, Pretzl GmbH

Tel.: +49 (0)172 617 8431

E-mail: birgit.schoeniger@pretzl.com

Strona www: www.pretzl.com