



Pressmeddelande

KIOXIA uppnår 4,8 miljarder högdimensionella vektorsökningsdatabaser på en enda server, varav 7.8x Index Build Time Acceleration via GPU:er

Utnyttjar NVIDIA cuVS-biblioteket med KIOXIA AiSAQ™-teknik för att indexera vektorer på 1024 dimensioner med minimal DRAM-användning

Tyskland, Düsseldorf, 16 mars 2026 – [KIOXIA Europe GmbH](#) tillkännagav idag KIOXIA Corporations framgångsrika demonstration av att uppnå högdimensionell vektorsökningsskalning till 4,8 miljarder vektorer på en enda server med sin öppna källkodsteknik KIOXIA AiSAQ™ approximate nearest neighbor search (ANNS). Dessutom visade KIOXIA en betydande minskning av indexbyggtiden genom att utnyttja GPU-acceleration via [NVIDIA cuVS](#). Dessa två framgångar markerar ett betydande framsteg för sökningslösningar för sökning inom återhämtning av förstärkt generering (RAG). Fortsatt utveckling pågår för att stödja större utplaceringar utöver 4,8 miljarder vektorer.

Indexbyggtid på en massiv vektordatabas är en avgörande utmaning för branschen. I samarbete med NVIDIA visade KIOXIA upp till 20-faldig förbättring i KIOXIA AiSAQ™-indexbyggtiden för högdimensionella vektorer med 1024 dimensioner, och upp till 7,8 gånger förbättring i end-to-end-byggtider. Denna 20-faldiga förbättring innebär en minskning från 28,4 dagar med CPU till 1,4 dagar med fyra [NVIDIA Hopper-GPU:er](#) för att bygga indexet, och en minskning från 31 dagar till 4 dagar i end-to-end-testning.¹



AI-applikationer kan nu förlita sig på större volymer vektoriserad information som når tiotals miljarder vektorer och mer därtill, lagrade på SSD-enheter, medan DRAM ensamt blir opraktisk även i miljardskala. KIOXIA möjliggör en mycket skalbar lagringsarkitektur med KIOXIA AiSAQi™-teknik genom att uppnå sökningar i miljardskala, vilket överträffar RAG-applikationslatenskraven med en enda frågeserver i en Milvus vectorDB-miljö driven av GPU-acceleration på indexbyggen som gör storskaliga installationer praktiska.

"Vektordatabaser utgör en ryggrad för applikationer som behöver förstå avsikt, kontext och likhet över massiva, ostrukturerade datamängder i realtid," säger Jason Hardy, Vice President för lagringsteknik på NVIDIA. "Genom att utnyttja GPU-accelererad indexering med NVIDIA cuVS-biblioteket stödjer KIOXIA högdimensionella vektordatabaser som kan skala och bygga index med en aldrig tidigare skådad effektivitet."

Den öppna källkodstekniken i KIOXIA AiSAQ™ som [presenterades förra året](#) hanterar RAG:s skalbarhetsutmaningar genom att möjliggöra vektorsökning direkt från SSD-enheter, med minskad DRAM-användning. KIOXIA AiSAQ™-tekniken ger hög skalbarhet, vilket gör den väl lämpad både för multi-tenant-miljöer och storskaliga monolitiska indexdistributioner. Tekniken utnyttjar en innovativ Global Index-algoritm som kombinerar hybridkluster och grafsökning för att leverera effektiv vektorsökning i extrem skala. Med flexibla finjusteringsalternativ för att balansera prestanda och högvolumvektorskalbarhet gör KIOXIA AiSAQ™-programvaran storskaliga installationer mer tillgängliga och enklare att expandera.

"Att uppnå skalbar vektorsökning i miljardskala på en enda server är en viktig milstolpe för branschen," säger Axel Stoermann, Chief Technology Officer & Vice President på KIOXIA Europe GmbH. "Med KIOXIA AiSAQ™ SSD-baserad vektorsökning och GPU-accelererad indexering med NVIDIA cuVS minskar vi indexbyggtiderna från veckor till dagar – ett avgörande steg mot att göra storskaliga RAG-system operativa."



KIOXIA är fortsatt engagerat när det gäller att utveckla lagringsdrivna AI-lösningar som stödjer intelligent databehandling i stor skala och fortsätter att utveckla KIOXIA AiSAQ™ mot biljonvektordistributioner.

Länk för att ladda ner KIOXIA AiSAQ™-programvaran med öppen källkod:
<https://github.com/kioxia-jp/aisaq-diskann>.

###

Noter:

1: Totalt bearbetades 19,66 TB vektordata för denna benchmark. Prestanda eller benchmark-resultat kan variera beroende på värdenheten, läs- och skrivvillkor, filstorlek och andra faktorer.

KIOXIA AiSAQ är ett varumärke som tillhör KIOXIA.

Företagsnamn, produktnamn och namn på tjänster kan vara varumärken som tillhör tredjepartsföretag.

Om KIOXIA Europe GmbH

KIOXIA Europe GmbH är det Europabaserade dotterbolaget till KIOXIA Corporation, en världsledande leverantör av flashminnen och SSD-enheter (Solid State Drives). Från uppfinningen av NAND-flashminnet till dagens berömda BiCS FLASH™ 3D-flashminne fortsätter KIOXIA att bana väg för innovativa minneslösningar och tjänster som berikar människors liv och vidgar samhällets horisonter. Företagets innovativa 3D-flashminnesteknik, BiCS FLASH™, formar framtidens lagringslösningar för applikationer med hög densitet såsom avancerade smarttelefoner, datorer, fordonssystem, datacenter och generativa AI-system.

Besök [KIOXIAs webbplats](#)

Kontaktuppgifter för publicering:

KIOXIA Europe GmbH, Hansaallee 183, 40549 Düsseldorf, Tyskland

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

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

KIOXIA

Kontaktuppgifter för redaktionella förfrågningar:

Lena Hoffmann, KIOXIA Europe GmbH

Tel: +49 (0) 211 36877 382

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

Utfärdat av:

Birgit Schöniger, Pretzl GmbH

Tel: +49 (0)172 617 8431

E-post: birgit.schoeniger@pretzl.com

Webb: www.pretzl.com