



Communiqué de presse

KIOXIA met en place une base de données de recherche vectorielle à haute dimension de 4,8 milliards d'entrées sur un seul serveur, avec un temps de création d'index 7,8fois plus rapide grâce aux GPU

Exploiter la bibliothèque NVIDIA cuVS avec la technologie KIOXIA AiSAQ™ pour indexer des vecteurs de 1024 dimensions avec une utilisation minimale de DRAM

Allemagne, Düsseldorf, 16 mars 2026 – [KIOXIA Europe GmbH](#) a annoncé aujourd'hui que KIOXIA Corporation avait réussi à démontrer la mise à l'échelle de la recherche vectorielle haute dimension à 4,8 milliards de vecteurs sur un seul serveur grâce à sa technologie open source KIOXIA AiSAQ™ de recherche approximative du plus proche voisin (ANNS). En outre, KIOXIA a démontré une réduction significative du temps de construction des index en tirant parti de l'accélération GPU via [NVIDIA cuVS](#). Ces deux avancées marquent un progrès important pour les solutions de recherche de type retrieval augmented generation (RAG). Le développement est en cours pour soutenir des déploiements à plus grande échelle au-delà de 4,8 milliards de vecteurs.

Le temps de construction des index dans les bases de données vectorielles à très grande échelle constitue un point critique majeur pour l'industrie. En collaboration avec NVIDIA, KIOXIA a démontré une amélioration allant jusqu'à 20 fois du temps de construction de l'indice KIOXIA AiSAQ™ pour les vecteurs de haute dimension de 1024 dimensions, et jusqu'à 7,8 fois d'amélioration des temps de construction de bout en bout. Cette amélioration de 20 fois représente une réduction de 28,4 jours avec un CPU à 1,4 jour avec quatre [GPU NVIDIA Hopper](#) pour créer l'index, et une réduction de 31 jours à 4 jours pour les tests de bout en bout.¹



Les applications d'IA peuvent désormais s'appuyer sur des volumes beaucoup plus importants de données vectorisées, atteignant des dizaines de milliards de vecteurs, voire davantage, stockés sur SSD, tandis que l'utilisation exclusive de la DRAM devient impraticable dès l'échelle du milliard. KIOXIA permet une architecture de stockage hautement évolutive grâce à la technologie KIOXIA AiSAQ™, en réalisant des recherches à l'échelle du milliard de vecteurs et en respectant les exigences de latence des applications RAG avec un seul serveur de requêtes, dans un environnement de base de données vectorielle Milvus, reposant sur l'accélération GPU pour la construction des index, rendant ainsi les déploiements à grande échelle réellement exploitables.

« Les bases de données vectorielles constituent une infrastructure essentielle pour les applications qui doivent comprendre l'intention, le contexte et les similarités au sein de vastes ensembles de données non structurées en temps réel », déclare Jason Hardy, vice-président des technologies de stockage chez NVIDIA. « En exploitant l'indexation accélérée par GPU avec la bibliothèque NVIDIA cuVS, KIOXIA prend en charge des bases de données vectorielles de haute dimension capables de monter en charge et de construire des index avec une efficacité sans précédent. »

[Annoncée pour la première fois l'année dernière](#), la technologie logicielle open source KIOXIA AiSAQ™ répond aux défis de scalabilité RAG en permettant la recherche vectorielle directement depuis les SSD, avec une utilisation réduite de DRAM. La technologie KIOXIA AiSAQ™ offre une grande évolutivité, ce qui la rend adaptée aussi bien aux environnements multi-locataires qu'aux déploiements d'index monolithiques à grande échelle. Elle s'appuie sur un algorithme innovant d'index global combinant un clustering hybride et une recherche par graphe afin de fournir une recherche vectorielle efficace à très grande échelle. Avec des options d'ajustement flexibles pour équilibrer performance et scalabilité vectorielle à haut volume, le logiciel KIOXIA AiSAQ™ rend les déploiements à grande échelle plus accessibles et faciles à développer.



« Atteindre une recherche vectorielle évolutive à l'échelle de plusieurs milliards de vecteurs sur un seul serveur constitue une avancée majeure pour l'industrie », déclare Axel Störmann, directeur technique (CTO) et vice-président de KIOXIA Europe GmbH. « Avec la recherche vectorielle basée sur SSD KIOXIA AiSAQ™ et l'indexation accélérée par GPU via NVIDIA cuVS, nous réduisons les temps de construction des index de plusieurs semaines à quelques jours. Il s'agit d'une étape essentielle pour rendre opérationnels les systèmes RAG à grande échelle. »

KIOXIA reste engagée dans le développement de solutions d'IA centrées sur le stockage, permettant un traitement intelligent des données à grande échelle, et continue de faire évoluer KIOXIA AiSAQ™ vers des déploiements atteignant des milliers de milliards de vecteurs.

Veuillez suivre le lien pour télécharger le logiciel open source KIOXIA AiSAQ : <https://github.com/kioxia-jp/aisaq-diskann>.

###

Remarques :

1 : Un total de 19,66 To de données vectorielles ont été traitées pour ce benchmark. Les performances ou résultats peuvent varier en fonction du système hôte, des conditions de lecture et d'écriture, des volumes de données et d'autres facteurs.

KIOXIA AiSAQ est une marque commerciale de KIOXIA.

Les noms de sociétés, de produits et de services mentionnés ici peuvent être des marques de leurs sociétés tierces.

À propos de KIOXIA Europe GmbH

KIOXIA Europe GmbH est la filiale européenne de KIOXIA Corporation, l'un des principaux fournisseurs mondiaux de mémoires flash et de disques SSD (Solid State Drive). De l'invention de la mémoire flash NAND à la célèbre mémoire flash 3D BiCS FLASH™ d'aujourd'hui, KIOXIA continue de proposer des solutions en matière de mémoire et des services innovants qui enrichissent la vie des gens et élargissent les horizons de la société. La technologie de mémoire flash 3D innovante BiCS FLASH™ de la société



façonne l'avenir du stockage dans les applications à haute densité, notamment les smartphones avancés, les PC, les systèmes automobiles, les centres de données et les systèmes d'IA générative.

Visitez notre [site Web KIOXIA](#)

Coordonnées pour la publication :

KIOXIA Europe GmbH, Hansaallee 183, 40549 Düsseldorf, Allemagne

Tél. : +49 (0)211 368 77-0

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

Coordonnées pour les demandes de nature rédactionnelles :

Lena Hoffmann, KIOXIA Europe GmbH

Tél. : +49 (0) 211 36877 382

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

Publié par :

Birgit Schöniger, Pretzl GmbH

Tél. : +49 (0)172 617 8431

E-mail : birgit.schoeniger@pretzl.com

Site Web : www.pretzl.com