



Пресс-релиз

КIOXIA достигает объема базы данных высокомерного векторного поиска в 4,8 млрд векторов на одном сервере, включая 7.8x Index Time Acceleration через GPU

Использование библиотеки NVIDIA cuVS с технологией KIOXIA AiSAQ™ для индексирования векторов 1024 измерений с минимальным использованием DRAM

Германия, Дюссельдорф, 16 марта 2026 года – [KIOXIA Europe ГмБХ](#) сегодня объявила об успешной демонстрации корпорации KIOXIA по достижению масштабирования векторного поиска с высокой размерностью до 4,8 миллиарда векторов на одном сервере с помощью открытой технологии KIOXIA AiSAQ™ Approximate Nearest Neighbor Search (ANNS). Кроме того, KIOXIA продемонстрировала значительное сокращение времени сборки индекса, используя ускорение GPU через [NVIDIA cuVS](#). Эти два достижения представляют собой значительный прогресс в поисковых решениях с дополненной генерацией поиска (RAG). Продолжается разработка для поддержки крупномасштабных развертываний свыше 4,8 миллиарда векторов.

Время наращивания индекса на массовой векторной базе данных — это важная проблема для отрасли. В сотрудничестве с NVIDIA корпорация KIOXIA продемонстрировала до 20-кратное улучшение времени сборки индекса KIOXIA AiSAQ™ для высокоразмерных векторов 1024 измерений и до 7,8-кратного повышения сквозного времени сборки. Это улучшение в 20 раз означает сокращение с 28,4 дн. использования процессора до 1,4 дн. при использовании четырёх [видеокарт NVIDIA Hopper](#) для построения индекса, а также сокращение с 31 дн. до 4 дн. при сквозном тестировании.¹



Приложения ИИ теперь могут опираться на большие объёмы векторной информации, охватывающей десятки миллиардов и более векторов, хранящихся на SSD, в то время как одна только DRAM становится непрактичной даже на миллиардном масштабе. KIOXIA обеспечивает высокомасштабируемую архитектуру хранения с технологией KIOXIA AiSAQ™, обеспечивая поиск в миллиард масштабов, превышая требования по задержке приложений RAG с помощью одного сервера запросов в среде Milvus vectorDB с ускорением GPU на индексных сборках, что делает масштабные развертывания практичными.

«Векторные базы данных служат основой для приложений, которым необходимо понимать намерение, контекст и сходство между огромными неструктурированными наборами данных в реальном времени, — сказал Джейсон Харди, вице-президент по технологиям хранения NVIDIA. — Используя ускоренное индексирование с помощью GPU с библиотекой NVIDIA cuVS, KIOXIA поддерживает крупноразмерные векторные базы данных, способные масштабироваться и создавать индексы с беспрецедентной эффективностью».

[Впервые анонсированная в прошлом году](#), открытая программная технология KIOXIA AiSAQ™ решает проблемы масштабируемости RAG, позволяя векторный поиск напрямую с SSD с уменьшением использования DRAM. Технология KIOXIA AiSAQ™ обеспечивает высокую масштабируемость, что делает её отлично подходящей как для сред с множеством арендаторов, так и для крупномасштабных монолитных индексных развертываний. Технология использует инновационный алгоритм Global Index, который сочетает гибридную кластеризацию и поиск по графам для обеспечения эффективного векторного поиска в экстремальных масштабах. Благодаря гибким настройкам для баланса производительности и масштабируемости векторов с высоким объёмом, программное обеспечение KIOXIA AiSAQ™ делает масштабные развертывания более доступными и проще для расширения.

«Достижение масштабируемого векторного поиска в многомиллиардном масштабе на одном сервере — это важная веха для отрасли», — сказал Аксель Стоерманн, технический & директор и вице-президент KIOXIA Europe GmbH.



«С помощью векторного поиска на базе KIOXIA AiSAQ™ SSD и ускоренного индексирования с помощью GPU с использованием NVIDIA cuVS мы сокращаем время сборки индекса с недель до дней — это критический шаг для обеспечения эксплуатации крупномасштабных RAG-систем».

KIOXIA по-прежнему стремится развивать решения на базе хранения данных, поддерживающие интеллектуальную обработку данных в масштабах, и продолжает развивать KIOXIA AiSAQ™ в направлении триллионных векторных развертываний.

Перейдите по ссылке, чтобы загрузить программное обеспечение с открытым исходным кодом KIOXIA AiSAQ: <https://github.com/kioxia-jp/aisaq-diskann>.

###

Примечания:

1: Для этого показателя было обработано всего 19,66 ТБ векторных данных. Скорость чтения и записи может варьироваться в зависимости от хост-устройства, условий чтения и записи и размера файла.

KIOXIA AiSAQ является товарным знаком компании KIOXIA.

Названия компаний, продуктов и услуг могут быть товарными знаками сторонних компаний.

О компании KIOXIA Europe GmbH

KIOXIA Europe GmbH — это европейское подразделение KIOXIA Corporation, ведущего мирового поставщика флэш-памяти и твердотельных накопителей (SSD). От изобретения флэш-памяти NAND до сегодняшней известной 3D-флэш-памяти BiCS FLASH™, KIOXIA продолжает внедрять передовые решения и услуги в области памяти, которые обогащают жизнь людей и расширяют горизонты общества. Инновационная технология 3D-флэш-памяти KIOXIA, BiCS FLASH™, определяет будущее хранения данных в ресурсоёмких приложениях, таких как передовые смартфоны, ПК, автомобильные системы, центры обработки данных и системы генеративного ИИ.

Посетите [сайт KIOXIA](#)

KIOXIA

Контактные данные для публикации:

KIOXIA Europe GmbH, Hansaallee 183, 40549 Дюссельдорф, Германия

Тел.: +49 (0)211 368 77-0

Электронная почта: KIE-support@eu.kioxia.com