

KALRAY LANCE NGENEA® POUR L'IA, UNE PLATEFORME D'ACCÉLÉRATION DES DONNÉES POUR L'IA GENERATIVE (GENAI) ET RAG

La solution Ngenea® pour l'IA offre des performances et une efficacité inégalées, en accélérant le traitement des données tout en simplifiant leur accès et en offrant un espace global unifié

Grenoble, France et Las Vegas, NV - 22 mai 2024 – Kalray (Euronext Growth Paris : ALKAL), leader dans les technologies matérielles et logicielles dédiées à la gestion et au traitement intensif des données du Cloud au Edge, annonce le lancement de « Ngenea® for AI », une nouvelle édition de sa plateforme d'accélération des données, spécialement optimisée pour les flux de données d'IA. Alors que l'IA exige de nouvelles façons de traiter et d'accéder à des volumes massifs de données, « Ngenea® for AI » aide les entreprises à accélérer leurs performances, à simplifier la gestion des données et à ouvrir leurs données à de nouveaux cas d'utilisation basés sur l'IA.

« Ngenea® for AI » est le complément des deux éditions de Ngenea® existantes qui adressent aujourd'hui les clients du monde des médias et du divertissement et du calcul haute performance (HPC). La nouvelle version de Ngenea® permet aux acteurs du monde de l'IA d'accélérer leurs performances dans le traitement de leurs données et d'accéder à leurs données non structurées à partir d'un espace global et unifié. Kalray a présenté au Dell Technologies World 2024 à Las Vegas sur son stand, Ngenea® for AI sur les serveurs Dell.

Le marché de l'IA générative (GenAI) devrait connaître [un taux de croissance annuel \(CAGR\) de 42%](#) au cours des 10 prochaines années, atteignant 1,3 trillion\$, contre seulement 40 milliards \$ en 2022 (source : Bloomberg Intelligence). Cependant, des applications basées, par exemple, sur de grands modèles de langage (LLM), des approches du type « Retrieval-Augmented Generation » (RAG), ou encore des applications comme « Smart Vision », mettent à l'épreuve la capacité des systèmes informatiques à déplacer des données vers et depuis leurs systèmes de stockage de données.

OUVRIR L'ACCÈS À L'IA GENERATIVE (GENAI) ET À L'APPROCHE « RAG »

« Ngenea® for AI » ajoute des capacités d'indexation et de recherche de données que les utilisateurs peuvent exploiter pour alimenter facilement n'importe quelle donnée à leurs applications à base d'intelligence artificielle (IA), comme les applications Smart Vision, GenAI et RAG. La nouvelle offre permet aux clients de l'IA de tirer pleinement parti de leurs actifs existants lors du déploiement d'applications à base d'IA quel que soit l'endroit où

les données sont stockées. « Ngenea for AI » simplifie et automatise ce qui serait autrement un flux de travail très complexe du point de vue de la gestion des données.

Jeff Necciai, Directeur technique de Duos Technologies, un des premiers utilisateurs de « Ngenea for AI » a déclaré : *"Le traitement rapide des données d'IA est crucial pour notre activité de détection des anomalies des wagons dont le but est d'identifier des problèmes potentiels qui pourraient conduire à des pannes ou des déraillements. Avec Ngenea® et les produits de Dell Technologies, Duos a obtenu une amélioration de performance de 120x sur notre solution Railcar Inspection Portal (RIP®), et une augmentation par 8 sur la précision des inspections. Travailler avec Kalray nous a aidé à placer les chemins de fer à l'avant-garde de l'innovation."*

UN NIVEAU DE STOCKAGE HAUTE PERFORMANCE POUR ALIMENTER LES GPU SANS EFFORT

La plateforme « Ngenea® for AI » intègre du stockage haute performance (Tier0) pour les charges de travail d'IA les plus intensives en données. Ce stockage Tier0 se base sur un système de fichiers parallèle haute performance éprouvé, capable de gérer facilement des pétaoctets de données et des milliards de fichiers – que les données soient en périphérie (« edge »), dans le cloud ou sur site. Ce stockage hyper performant permet aux clients de NGenea® de maximiser en particulier leur investissement dans les GPU, en optimisant leur utilisation.

Avec Ngenea®, les développeurs d'IA peuvent :

- **Accélérer les flux de données liés à l'IA** et améliorer la vitesse de traitement de manière globale des opérations d'IA.
- **Améliorer les performances des modèles d'IA** en alimentant les GPU plus rapidement, permettant des calculs plus rapides
- **Bénéficier de capacités d'indexation avancées** qui permettent de fournir des contextes plus pertinents aux applications d'IA générative et RAG, et de permettre ainsi de répondre aux requêtes de manière plus efficaces et plus précises
- **Faciliter l'intégration et l'utilisation des données** de l'entreprise avec un ensemble grandissant d'outils utilisant la puissance de l'IA, permettant ainsi aux entreprises de valoriser leurs données par rapport aux besoins de leur business.

Les DPU de Kalray, avec en particulier la carte d'accélération TURBOCARD4 (TC4), offre l'option aux clients de « Ngenea for AI » d'accélérer encore plus le traitement des données pour les flux de travail ultra-exigeants. Les DPU ont une architecture complémentaire aux GPU et permettent d'en optimiser l'usage.

"Même avec les investissements considérables que les dirigeants sont prêts à faire dans le domaine de l'IA, très peu de CIO disposent des budgets nécessaires pour déployer un nouveau système uniquement dédié à l'IA," a déclaré Scott Sinclair, Directeur de pratique, infrastructure et modernisation à Enterprise Strategy Group™. "Des cas d'utilisation innovants qui permettent, aux entreprises de tirer parti de leur infrastructure existante et de déployer de nouvelles applications d'IA, adressent un besoin important sur le marché, inexploité à ce jour."

UN KIT D'OUTILS D'IA POUR UNE INTÉGRATION FACILITÉE

« Ngenea® for AI » inclut un kit (« SDK ») IA que les partenaires de Kalray peuvent utiliser pour une intégration plus facile de leurs applications avec Ngenea®. Par exemple, grâce à l'intégration de Ngenea® avec « Perifery Vision AI », les utilisateurs de Ngenea® peuvent rechercher intuitivement du contenu sur une vaste quantité de données.

"Les centres de données traditionnels n'ont pas été conçus pour permettre d'accéder facilement à tout l'ensemble de données massifs nécessaires pour les charges de travail dédiées à l'IA – ni pour traiter ces charges de travail rapidement et efficacement, » a déclaré Eric Baissus, Président du Directoire de Kalray. « Ngenea® for IA » est la plateforme d'accélération des données pour l'ère de l'IA – une solution qui augmente les performances tout en conservant les serveurs et les systèmes de stockage existants et qui permet, grâce à notre toolkit, de simplifier l'intégration des applications d'IA. Avec notre logiciel, les entreprises et les fournisseurs de services peuvent accélérer l'utilisation de leur GPU et leur intégration de l'IA de manière générale pour que l'IA offre véritablement un avantage concurrentiel."

La plateforme « Ngenea® for AI » rejoint les deux autres éditions de Ngenea® qui simplifient l'intégration des applications pour des cas d'utilisation spécifiques : « Ngenea® for HPC » – pour le monde du « High Computing » (HCP), et « Ngenea® for M&E » pour accélérer les flux de travail dans le monde des médias. « Ngenea® for IA » est disponible dès aujourd'hui chez Dell, Kalray et certains partenaires.

Pour plus d'informations, cliquez [ici](#).

À PROPOS DE KALRAY

Kalray (Euronext Growth Paris : ALKAL) est leader dans les technologies matérielles et logicielles dédiées à la gestion et au traitement intensif des données du Cloud au Edge. Kalray propose une gamme complète de produits et de solutions permettant de développer des applications, ainsi que des infrastructures, plus intelligentes, plus efficaces et plus économes en énergie.

L'offre de Kalray comprend des processeurs DPU (Data Processing Unit) brevetés, issus d'une technologie unique, des cartes d'accélération, ainsi que des solutions logicielles innovantes de stockage et de gestion des données. Utilisées séparément ou de façon combinée, les solutions hautes performances de Kalray permettent à ses clients d'améliorer l'efficacité des datacenters ou de concevoir les meilleures solutions dans des secteurs à forte croissance tels que l'IA, les médias et le divertissement, les sciences de la vie, la recherche scientifique, l'Edge Computing, l'automobile, et bien d'autres.

Essaimage du CEA fondé en 2008, Kalray compte parmi ses investisseurs industriels et financiers Alliance Venture (Renault-Nissan-Mitsubishi), NXP Semiconductors, Bpifrance. Kalray, c'est la technologie, l'expertise, la passion d'offrir toujours plus : plus pour un monde intelligent, plus pour la planète, plus pour nos clients et les développeurs. www.kalrayinc.com

CONTACTS INVESTISSEURS

Eric BAISSUS

contactinvestisseurs@kalrayinc.com

Tel. +33 4 76 18 90 71

ACTUS Finance & Communication

Anne-Pauline PETUREAUX

kalray@actus.fr

Tel. + 33 1 53 67 36 72

CONTACTS PRESSE

Penta STANLEY

communication@kalrayinc.com

Tel. +44 7939 877 880

Mary Kae Marinac

Contact presse Kalray

mk@mkmarinac.com

Tel. +1 978-685-3136

ACTUS Finance & Communication

Serena BONI

sboni@actus.fr

Tel. +33 4 72 18 04 92