



Nvidia lancia Vera, il primo processore pensato per gli agenti IA

Descrizione

(Adnkronos) Il settore tecnologico globale si prepara a una transizione infrastrutturale significativa con il lancio di NVIDIA Vera, una nuova classe di processori ottimizzati per l'intelligenza artificiale agentica. A differenza delle architetture general-purpose, Vera promette di accelerare il completamento delle attività di 1,8 volte, ponendo le basi per una nuova economia dei data center focalizzata sulla generazione di token piuttosto che sulla mera densità dei core. Il processore è il risultato di un percorso iniziato con le CPU Grace, che hanno già raggiunto una diffusione capillare con circa 2,5 milioni di unità distribuite, consolidando la posizione di NVIDIA nel segmento dei data center. L'adozione di Vera è trasversale e coinvolge attori di primo piano in ambiti differenti. Dalla finanza, con il NYSE che punta a ottimizzare la latenza della propria infrastruttura, a laboratori di ricerca come Anthropic e OpenAI, fino ai grandi hyperscaler come Oracle Cloud Infrastructure e ByteDance. Anche i principali produttori di sistemi, tra cui Dell, HPE, Lenovo e Supermicro, hanno già integrato la tecnologia nei propri roadmap hardware. L'obiettivo comune è supportare una nuova generazione di AI in grado non solo di elaborare risposte, ma di eseguire codice autonomamente, utilizzare strumenti esterni e valutare i risultati delle proprie azioni. Dal punto di vista tecnico, il cuore del sistema è rappresentato dai core Olympus, progettati per gestire in modo efficiente i runtime Python, l'esecuzione di codice in ambiente sandbox e la logica di orchestrazione. Con 88 core, la tecnologia Spatial Multithreading e un sottosistema di memoria LPDDR5X capace di 1,2 TB/s di banda, Vera è pensata per eliminare i colli di bottiglia computazionali che spesso rallentano le pipeline di dati. Questo permette di mantenere gli acceleratori costantemente alimentati, migliorando drasticamente il throughput complessivo delle cosiddette AI factory. L'integrazione nell'ecosistema esistente è garantita dalla tecnologia di interconnessione NVLink-C2C di seconda generazione, che permette una comunicazione coerente con le GPU fino a 1,8 TB/s. Inoltre, il processore Vera BlueField-4 STX amplia le capacità includendo accelerazione di rete e storage, oltre a funzionalità di sicurezza integrata. Con una disponibilità prevista a partire dal prossimo autunno, il sistema si propone come la prima alternativa concreta alle architetture x86 per il segmento dei server CPU standalone, offrendo soluzioni sia in configurazioni raffreddate a liquido per grandi cluster, sia in sistemi ad aria per implementazioni enterprise.

tecnologiawebinfo@adnkronos.com (Web Info)

Categoria

1. News Nazionali

Tag

- 1. adnkronos
- 2. tecnologia

Data di creazione

Giugno 1, 2026

Autore

admin

default watermark