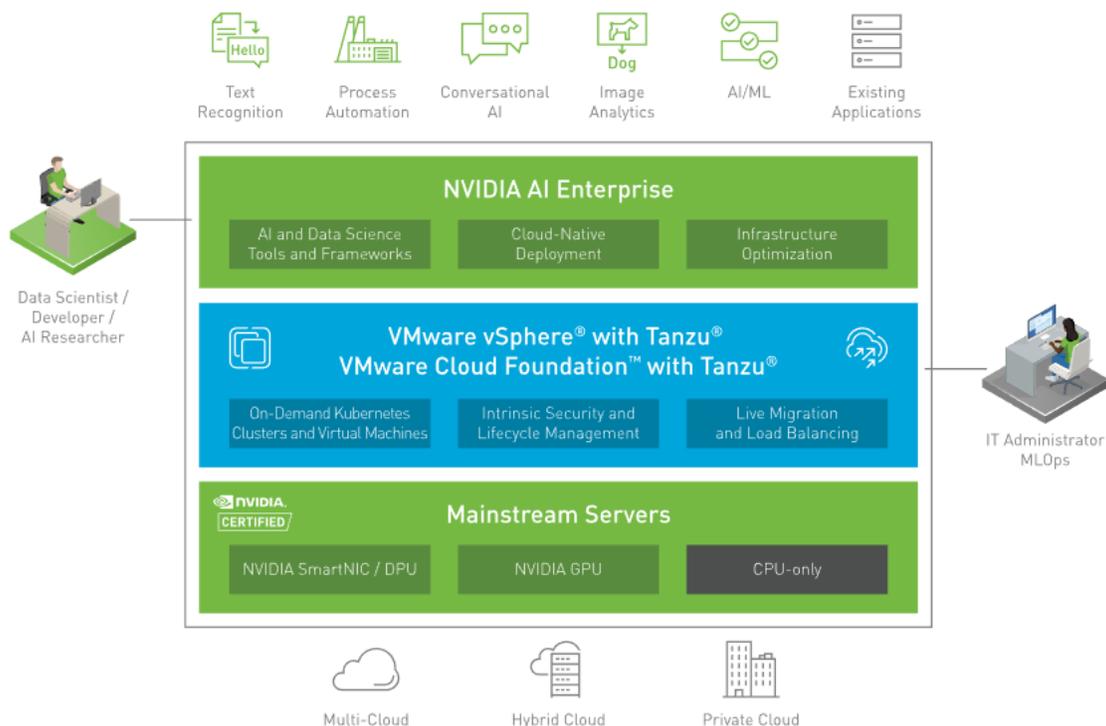


Nvidia AI Enterprise for VMware

Nvidia AI Enterprise è una soluzione software AI end-to-end, cloud-native per l'analisi dei dati. È ottimizzato, certificato e supportato da Nvidia per essere eseguito su data center virtualizzati insieme a VMware vSphere con Tanzu e VMware Cloud Foundation con Tanzu su sistemi certificati Nvidia.

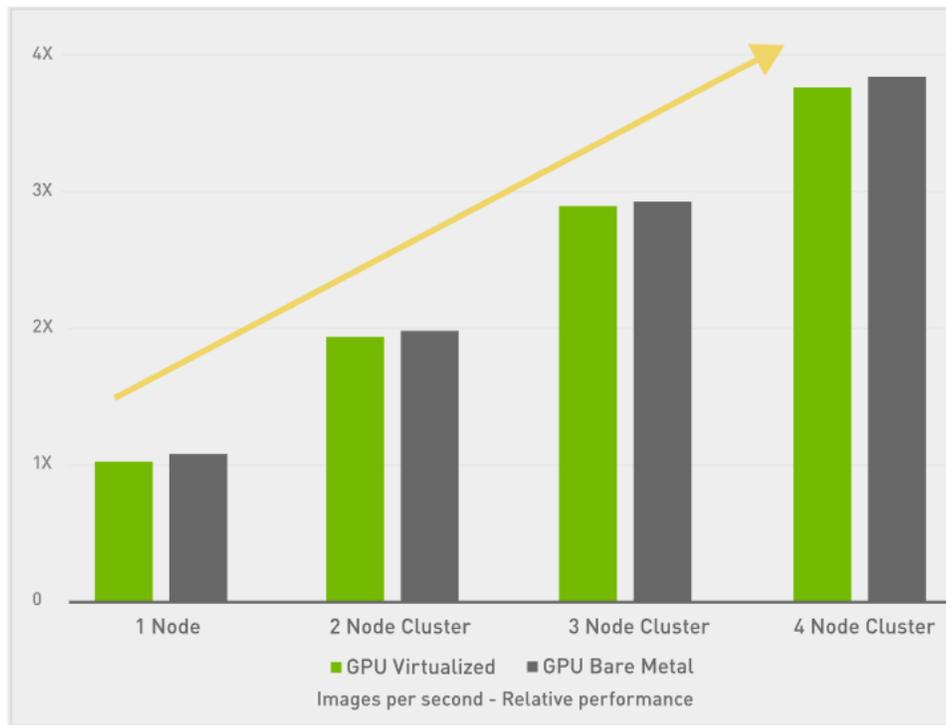
Piattaforma AI-Ready VMware + Nvidia



La suite Nvidia AI Enterprise include le applicazioni, i framework e gli strumenti usati dall'AI per i ricercatori, i data scientists e gli sviluppatori.

VMware vSphere e VMware Cloud Foundation offre alla flessibilità nel eseguire workloads AI direttamente sulle macchine virtuali o sui container in esecuzione sulle stesse macchine virtuali. Il ruolo di VMware Tanzu for container è proprio quello di orchestrare e gestire dei cluster Kubernetes.

È anche importante evidenziare come la virtualizzazione supportata da VMware fornisca le stesse performance del Bare-Metal per AI, si noti infatti come la suite, nel seguente diagramma, fornisca un livello quasi pari di performance tra bare-metal e GPU virtualizzata:



VMware vSphere with Tanzu e VMware Tanzu Kubernetes Grid services uniti a Nvidia AI Enterprise rendono accessibili a VMware Tanzu l'accesso alle GPU e DPU Nvidia per accelerare il containerized AI e i workloads per Machine Learning.

Usando vCenter, i GPU accelerated hosts possono essere facilmente gestiti e mantenuti per frazionare le risorse computazionali nell'ambiente Tanzu.

VMware vSphere with Tanzu e VMware Cloud Foundation with Tanzu sono i leader del settore per quanto riguarda la virtualizzazione della piattaforma che ha in esecuzione oltre 70 milioni di workloads tra centinaia di migliaia di organizzazioni in tutto il mondo.

Si occupa di trasformare i server bare-metal (includendo risorse CPU e GPU) in AI centralizzata e pools di infrastrutture gestite in maniera centralizzata che possono effettuare velocemente il provisioning multiplo di Kubernetes accelerato usando Tanzu. VMware vSphere ottimizza le prestazioni, aumenta l'affidabilità, rafforza la sicurezza, automatizza la manutenzione e riduce i costi per creare un'infrastruttura che supporta i workloads esistenti e le applicazioni di nuova generazione come l'AI.