

Il duplice ruolo dell'AI nei data center on-premise

Molte aziende oggi desiderano sfruttare la potenza dell'intelligenza artificiale e i data center rappresentano una piattaforma capace di supportare queste ambizioni. Tuttavia, per ottenere prestazioni ottimali è necessario disporre di un'infrastruttura e di un'architettura di rete ben progettate. L'AI migliora al contempo la gestibilità della rete e l'esperienza all'utente finale, ottenendo così un duplice impatto

Davide Bianchini

I processi aziendali fanno grande affidamento sulle applicazioni e, dunque, il loro corretto funzionamento è una priorità chiave per i chief information officer (CIO), specialmente nelle aziende che integrano in modo più completo l'intelligenza artificiale nelle proprie procedure. Queste applicazioni risiedono in genere nei data center delle aziende (on-premise), il che evidenzia la necessità di un ambiente di rete robusto e di un'architettura a esso associata altrettanto solida. In assenza di questi elementi, le applicazioni aziendali non possono funzionare in modo ottimale e il collegamento ai dispositivi aziendali è impossibile.

L'automazione è inevitabile

La gestione dell'ambiente di rete nei data center è diventata, negli ultimi anni, sempre più complessa. Una complessità acuita dalle diverse tecnologie e applicazioni in uso, spesso gestite da team distinti, che possono portare a problemi di comunicazione e iso-

lamento dei team di delivery. Senza contare che l'elevato volume di dati generati all'interno dei data center complica ulteriormente la rapida identificazione e analisi delle informazioni raccolte dagli strumenti di management. Il monitoraggio manuale, insieme ad attività come la configurazione, l'applicazione di patch e la risoluzione degli incidenti, è in genere dispendioso in termini di manodopera, inefficiente e soggetto a errori umani. Allo stesso modo, il monitoraggio e l'ottimizzazione manuale delle prestazioni della rete richiedono molto tempo. Inoltre, l'individuazione dei rischi per la sicurezza e delle potenziali anomalie diventa ancora più difficile senza strumenti avanzati di monitoraggio e analisi. Da ultimo, la manutenzione manuale tende a essere di natura reattiva, il che può comportare costi innecessari e interruzioni impreviste.

L'intelligenza artificiale offre una soluzione

L'automazione della gestione del data center offre invece numerosi vantaggi, tra cui la riduzione degli errori e la velocizzazione delle attività di routine. AIOps svolge un ruolo cruciale nel portare la gestione automatizzata a uno standard più elevato: integrando dati provenienti da varie fonti, fornisce una panoramica centralizzata e rivela schemi e correlazioni che spesso sfuggono agli amministratori umani. Ciò favorisce una migliore collaborazione tra gli amministratori, accelera la risoluzione dei problemi e offre una visione più olistica dell'ambiente di rete. Inoltre, il filtraggio e l'analisi automatica dei dati garantiscono che solo le informazioni più importanti e rilevanti siano rese disponibili agli amministratori. Ciò si traduce in meno lavoro per gli amministratori IT, analisi dei dati più efficienti e processi decisionali più rapidi ed efficaci. AIOps monitora continuamente il sistema e suggerisce ottimizzazioni per le attività di gestione della rete che tradizionalmente vengono eseguite ma-

Immagine di Vecstock su Freepik



La gestione dell'ambiente di rete nei data center è diventata, negli ultimi anni, sempre più complessa

nualmente: rileva gli incidenti in tempo reale, analizza le potenziali cause e consiglia o implementa proattivamente le soluzioni. In aggiunta, l'analisi predittiva consente ad AIOps di anticipare i problemi prima che si verifichino e consigliare gli interventi di manutenzione in base alle necessità reali, rendendo la gestione da reattiva a proattiva o addirittura preventiva. Dunque, AIOps offre un'esperienza di livello superiore per la gestione automatizzata dei data center, consentendo ad amministratori e dipendenti di godere di benefici eccezionali e concentrarsi su attività critiche, quali il miglioramento del valore delle applicazioni e l'utilizzo efficace dei dati.

Data center e carichi di lavoro AI

AIOps non solo migliora la gestione automatizzata e proattiva dei data center, ma li rende anche le piattaforme ideali per l'esecuzione di workload AI e l'elaborazione dei dati essenziali. Pertanto, è fondamentale configurare questi ambienti in modo ottimale, specialmente l'infrastruttura di rete, al fine di realizzare l'ambiente ideale per uno sviluppo proficuo delle tecnologie AI. Tuttavia, rispetto ai data center tradizionali, l'elaborazione dei workload di intelligenza artificiale presenta nuove sfide, principalmente dovute alle esigenze specifiche imposte dal traffico AI. Il cluster di computer, e soprattutto i processori GPU (che si trovano nei data center e sono alla base dell'AI), richiedono velocità elevate per l'elaborazione dei dati. Parimenti la larghezza di banda deve garantire capacità notevoli e costanti, raggiungendo velocità di 800 Gbps e oltre. In più, una scalabilità di questo tipo è fondamentale per soddisfare le diverse esigenze aziendali. Ciò implica l'utilizzo efficiente delle risorse del data center per ridurre i costi, garantire prestazioni costanti, prevenire sovraccarichi, allocare risorse ad attività critiche e ottimizzare il consumo energetico. La scalabilità consente inoltre ai data center di adattarsi rapidamente ai cambiamenti delle esigenze aziendali, migliorando la ridondanza e le attività di manutenzione con interruzioni minime delle attività.

Il giusto ambiente di rete per l'intelligenza artificiale

L'esecuzione di carichi di lavoro AI nei data center richiede un ambiente di rete idoneo, ad esempio grazie alla gestione del traffico AI tramite il protocollo Ethernet, anziché il tradizionale e sempre



L'intelligenza artificiale sta riscontrando un utilizzo diffuso in molte organizzazioni, modellandone sempre più le applicazioni e l'elaborazione dei dati, con i data center che svolgono un ruolo fondamentale

più di nicchia InfiniBand. L'utilizzo dell'intent-based networking, in aggiunta ad AIOps, renderà i data center molto più adatti all'elaborazione dei workload AI. L'architettura di rete intent-based consente alle aziende di definire gli obiettivi aziendali, dopodiché l'ambiente di rete del data center li converte automaticamente nelle giuste configurazioni, in base alle informazioni sulla rete, ai dati analizzati e all'orchestrazione. Queste configurazioni vengono continuamente convalidate tramite un meccanismo closed-loop al fine di controllare che gli obiettivi iniziali siano sempre soddisfatti e riadattarli automaticamente se necessario.

Il networking intent-based garantisce che i data center mantengano la corretta configurazione di rete e li ottimizza appositamente per la gestione dei carichi di lavoro AI. Questo approccio automatizzato non solo solleva gli amministratori dalle attività manuali, ma migliora anche costantemente l'esperienza dell'utente per soddisfare gli standard desiderati. Gli operatori di data center che utilizzano reti basate su intenti e AIOps per semplificare e perfezionare la progettazione, l'implementazione e lo svolgimento delle operazioni dell'infrastruttura potranno così differenziarsi sempre più dalla concorrenza.

Una piattaforma di networking AI-Native completa

L'intelligenza artificiale sta riscontrando un utilizzo diffuso in molte organizzazioni, modellandone sempre più le applicazioni e l'elaborazione dei dati, con i data center che svolgono un ruolo fondamentale. Un ambiente di rete ottimizzato per l'intelligenza artificiale garantisce che i workload AI possano essere elaborati in modo efficiente. Una piattaforma di rete AI nativa consente alle aziende e agli amministratori di gestire gli ambienti di lavoro tramite AIOps da un unico ambiente ed elaborare in modo ottimale i carichi di lavoro AI in condizioni controllate. Ciò permette ad aziende, amministratori e utenti finali di concentrarsi interamente sui propri obiettivi senza doversi preoccupare dell'infrastruttura e dell'architettura di rete richieste.

Juniper Networks - www.juniper.net



La manutenzione manuale tende a essere di natura reattiva, il che può comportare costi innessari e interruzioni impreviste