Ai Inferencing at the Edge - NetApp con Lenovo ThinkSystem - progettazione di soluzioni

NetApp Solutions

NetApp
April 26, 2024

This PDF was generated from https://docs.netapp.com/it-it/netapp-solutions/ai/ai-edge-introduction.html on April 26, 2024. Always check docs.netapp.com for the latest.

Sommario

Ai Inferencing at the Edge - NetApp con Lenovo ThinkSystem - progettazione di soluzioni 1
TR-4886: Ai Inferencing at the Edge - NetApp con Lenovo ThinkSystem - progettazione di soluzioni 1
Panoramica della tecnologia 6
Piano di test 13
Eseguire il test della configurazione 14
Procedura di test 16
Risultati del test 17
Opzioni di dimensionamento dell’architettura 22
Conclusione 23

Ai Inferencing at the Edge - NetApp con Lenovo ThinkSystem - progettazione di soluzioni

TR-4886: Ai Inferencing at the Edge - NetApp con Lenovo ThinkSystem - progettazione di soluzioni

Sathish Thyagarajan, NetApp Miroslav Hodak, Lenovo

Riepilogo

Diversi scenari applicativi emergenti, come i sistemi avanzati di assistenza alla guida (ADAS), Industry 4.0, smart cities e Internet of Things (IoT), richiedono l'elaborazione di flussi di dati continui con una latenza quasi nulla. Questo documento descrive un'architettura di calcolo e storage per implementare deduzione di intelligenza artificiale (ai) basata su GPU su storage controller NetApp e server Lenovo ThinkSystem in un ambiente edge che soddisfi questi requisiti. Questo documento fornisce inoltre dati sulle performance per il benchmark MLPerf Inference standard di settore, valutando varie attività di inferenza su edge server dotati di GPU NVIDIA T4. Analizziamo le performance degli scenari di inferenza offline, single stream e multistream e mostriamo che l'architettura con un sistema di storage condiviso in rete a costi contenuti è altamente performante e fornisce un punto centrale per la gestione di dati e modelli per più edge server.

Introduzione

Le aziende stanno generando sempre più enormi volumi di dati all'edge della rete. Per ottenere il massimo valore dai sensori intelligenti e dai dati IoT, le organizzazioni sono alla ricerca di una soluzione di streaming di eventi in tempo reale che consenta l'edge computing. I lavori più impegnativi dal punto di vista computazionale vengono quindi eseguiti sempre più all'edge, al di fuori dei data center. L'inferenza dell'ai è uno dei fattori trainanti di questa tendenza. Gli edge server forniscono una potenza di calcolo sufficiente per questi carichi di lavoro, soprattutto quando si utilizzano acceleratori, ma lo storage limitato è spesso un problema, soprattutto negli ambienti multisserver. In questo documento mostreremo come puoi implementare un sistema storage condiviso nell'ambiente edge e in che modo esso benefici per i carichi di lavoro di inferenza ai senza imporre penalizzazioni in termini di performance.

Questo documento descrive un'architettura di riferimento per l'inferenza ai ai edge. Combina più edge server Lenovo ThinkSystem con un sistema storage NetApp per creare una soluzione semplice da implementare e gestire. Si tratta di una guida di riferimento per implementazioni pratiche in varie situazioni, come ad esempio il reparto di produzione con telecamere multiple e sensori industriali, sistemi POS (Point of sale) nelle transazioni al dettaglio o sistemi FSD (Full Self-Driving) che identificano anomalie visive nei veicoli autonomi.

Il presente documento illustra il test e la convalida di una configurazione di calcolo e storage costituita da un server edge Lenovo ThinkSystem SE350 e da un sistema storage entry-level NetApp AFF ed EF-Series. Le architetture di riferimento offrono una soluzione efficiente e conveniente per le implementazioni ai, fornendo al contempo servizi dati completi, protezione integrata dei dati, scalabilità perfetta e storage dei dati connesso al cloud con NetApp ONTAP e il software di gestione dei dati NetApp SANtricity.

Pubblico di riferimento

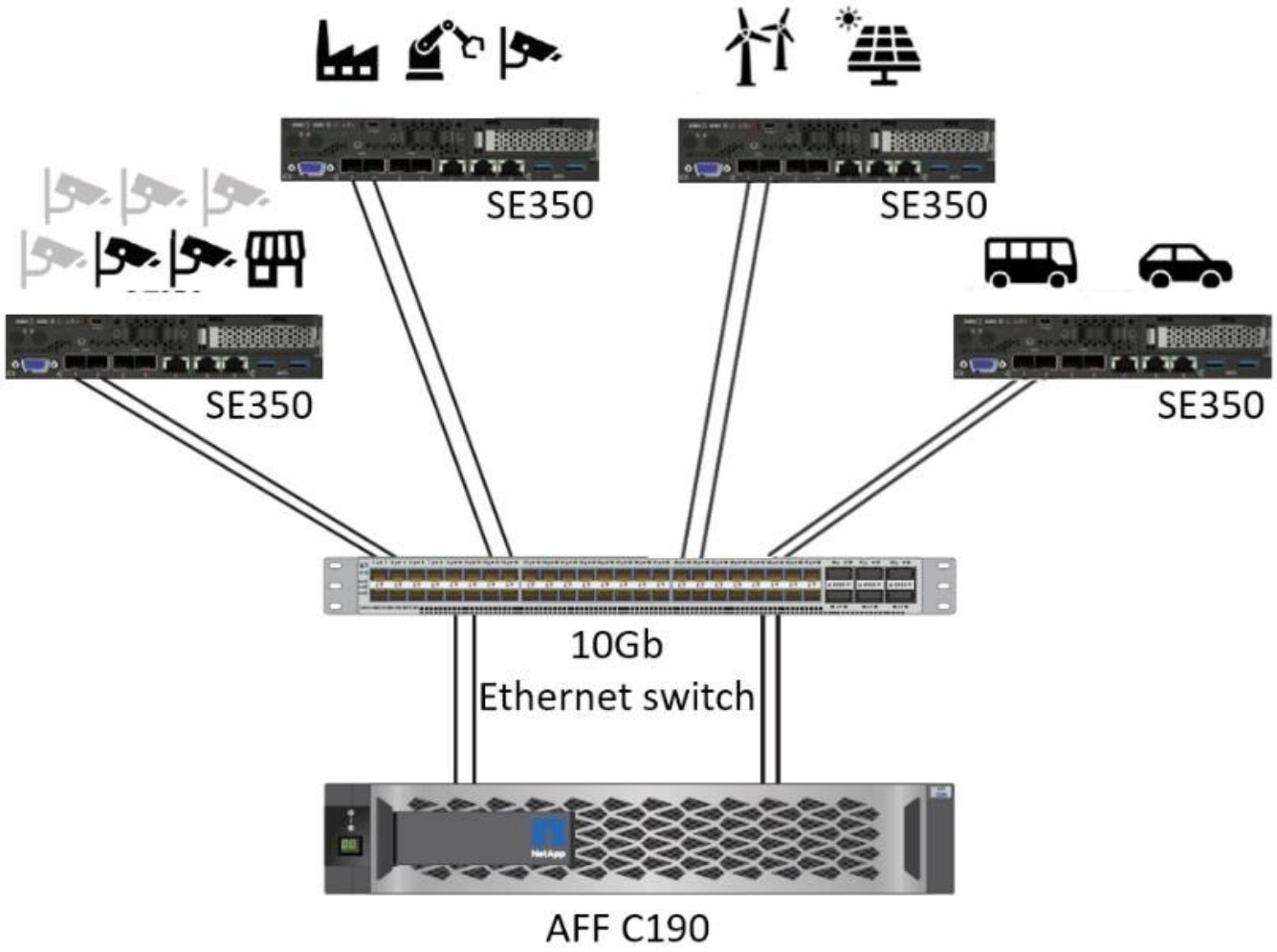
Il presente documento è destinato ai seguenti destinatari:

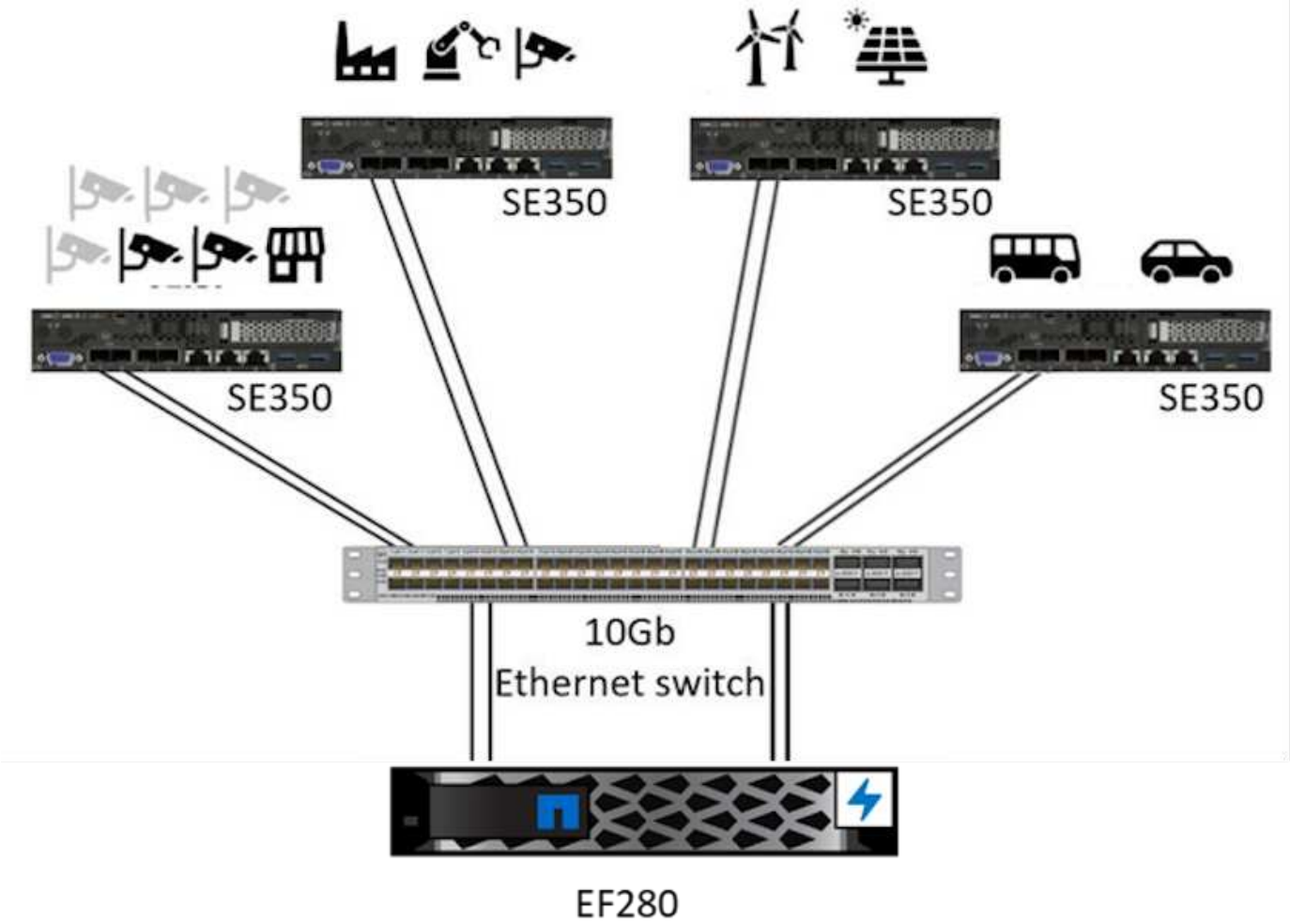
- Business leader e Enterprise architect che vogliono produrre l'ai alla periferia della rete.
- Data scientist, data engineer, ricercatori ai/machine learning (ML) e sviluppatori di sistemi ai.

- Architetti aziendali che progettano soluzioni per lo sviluppo di modelli e applicazioni ai/ML.
- Data scientist e ingegneri ai alla ricerca di modi efficienti per implementare modelli di deep learning (DL) e ML.
- Edge Device Manager e amministratori degli edge server responsabili dell'implementazione e della gestione dei modelli di inferenza edge.

Architettura della soluzione

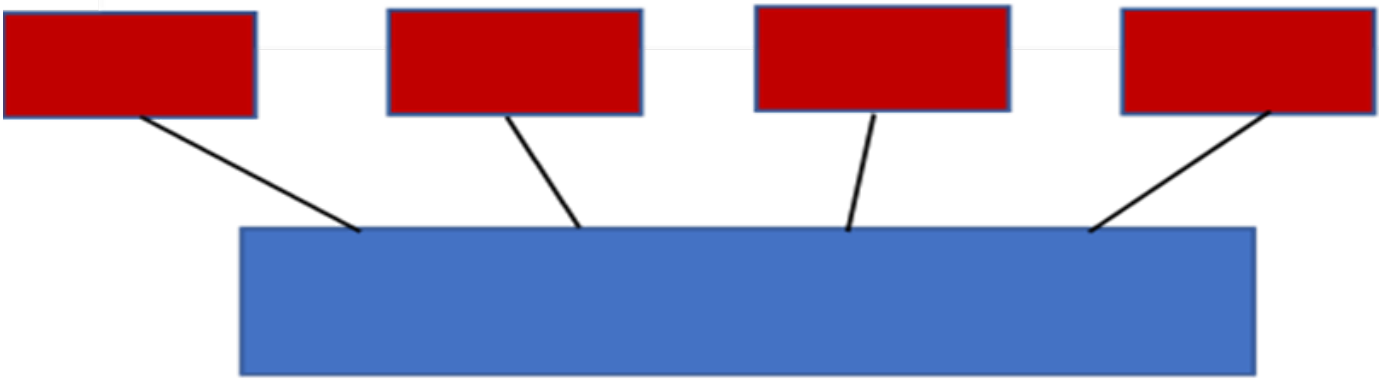
Questo server Lenovo ThinkSystem e la soluzione di storage NetApp ONTAP o NetApp SANtricity sono progettati per gestire l'inferenza ai su set di dati di grandi dimensioni utilizzando la potenza di elaborazione delle GPU insieme alle CPU tradizionali. Questa convalida dimostra performance elevate e una gestione ottimale dei dati con un'architettura che utilizza uno o più edge server Lenovo SR350 interconnessi con un singolo sistema di storage NetApp AFF, come mostrato nelle due figure seguenti.





La panoramica dell'architettura logica nella figura seguente mostra i ruoli degli elementi di calcolo e storage in questa architettura. In particolare, viene mostrato quanto segue:

- I dispositivi di calcolo edge che eseguono l'inferenza sui dati ricevuti da telecamere, sensori e così via.
- Un elemento storage condiviso che serve più scopi:
  - Fornisce una posizione centrale per i modelli di inferenza e altri dati necessari per eseguire l'inferenza. I server di calcolo accedono direttamente allo storage e utilizzano modelli di inferenza in tutta la rete senza la necessità di copiarli localmente.
  - I modelli aggiornati vengono inviati qui.
  - Archivia i dati di input ricevuti dagli edge server per un'analisi successiva. Ad esempio, se i dispositivi edge sono collegati alle telecamere, l'elemento di storage mantiene i video acquisiti dalle telecamere.



rosso	blu
Sistema di calcolo Lenovo	Sistema storage NetApp AFF
Dispositivi edge che eseguono deduzione sugli input provenienti da telecamere, sensori e così via.	Storage condiviso con modelli di inferenza e dati provenienti da dispositivi edge per analisi successive.

Questa soluzione NetApp e Lenovo offre i seguenti vantaggi principali:

- La GPU ha accelerato il computing all'edge della rete.
- Implementazione di più edge server supportata e gestita da uno storage condiviso.
- Protezione dei dati efficace per soddisfare gli obiettivi RPO (Recovery Point Objective) e RTO (Recovery Time Objective) ridotti senza perdita di dati.
- Gestione dei dati ottimizzata con copie Snapshot e cloni NetApp per ottimizzare i flussi di lavoro di sviluppo.

Come utilizzare questa architettura

Questo documento convalida la progettazione e le performance dell'architettura proposta. Tuttavia, non abbiamo testato alcuni componenti a livello di software, come la gestione di container, workload o modelli e la sincronizzazione dei dati con il cloud o il data center on-premise, perché sono specifici di uno scenario di implementazione. In questo caso, esistono diverse scelte.

A livello di gestione dei container, Kubernetes Container Management è una buona scelta ed è ben supportato sia in una versione completamente upstream (Canonical) che in una versione modificata adatta per le implementazioni Enterprise (Red Hat). Il "Piano di controllo ai di NetApp" Che utilizza NetApp Trident e il nuovo aggiunto "NetApp DataOps Toolkit" Offre tracciabilità integrata, funzioni di gestione dei dati, interfacce e strumenti per data scientist e data engineer da integrare con lo storage NetApp. Kubeflow, IL toolkit ML per Kubernetes, offre funzionalità ai aggiuntive, oltre al supporto per il controllo delle versioni dei modelli e KFServing su diverse piattaforme, come TensorFlow Serving o NVIDIA Triton Inference Server. Un'altra opzione è la piattaforma NVIDIA EGX, che offre la gestione dei carichi di lavoro e l'accesso a un catalogo di container di inferenza ai abilitati per GPU. Tuttavia, queste opzioni potrebbero richiedere sforzi ed esperienza significativi per metterle in produzione e potrebbero richiedere l'assistenza di un vendor di software indipendente (ISV) o di un consulente di terze parti.

Aree di soluzione

Il vantaggio principale dell'inferenza ai e dell'edge computing è la capacità dei dispositivi di calcolare, elaborare e analizzare i dati con un elevato livello di qualità senza latenza. Esistono troppi esempi di casi d'utilizzo di edge computing da descrivere in questo documento, ma di seguito sono riportati alcuni esempi importanti:

Automobili: Veicoli autonomi

La classica illustrazione dell'edge computing si trova nei sistemi avanzati di assistenza alla guida (ADAS) nei veicoli autonomi (AV). L'ai nelle auto senza conducente deve elaborare rapidamente una grande quantità di dati provenienti da telecamere e sensori per essere un pilota sicuro e di successo. Un'interpretazione troppo lunga tra un oggetto e un essere umano può significare vita o morte, pertanto è fondamentale essere in grado di elaborare i dati il più vicino possibile al veicolo. In questo caso, uno o più server di calcolo edge gestiscono l'input da telecamere, RADAR, LDAR e altri sensori, mentre lo storage condiviso contiene modelli di inferenza e memorizza i dati di input provenienti dai sensori.

Settore sanitario: Monitoraggio dei pazienti

Uno dei maggiori impatti dell'intelligenza artificiale e dell'edge computing è la sua capacità di migliorare il monitoraggio continuo dei pazienti per le malattie croniche sia nelle strutture di assistenza domiciliare che nelle unità di terapia intensiva (ICU). I dati provenienti da dispositivi periferici che monitorano i livelli di insulina, la respirazione, l'attività neurologica, il ritmo cardiaco e le funzioni gastrointestinali richiedono un'analisi istantanea dei dati che devono essere utilizzati immediatamente, in quanto il tempo necessario per agire è limitato per salvare la vita di qualcuno.

Retail: Pagamento senza cassa

L'edge computing può potenziare ai e ML per aiutare i retailer a ridurre i tempi di checkout e aumentare il traffico. I sistemi senza cassiere supportano diversi componenti, ad esempio:

- Autenticazione e accesso. Collegare l'acquirente fisico a un account validato e consentire l'accesso allo spazio di vendita al dettaglio.
- Monitoraggio dell'inventario. Utilizzo di sensori, tag RFID e sistemi di visione computerizzata per confermare la selezione o la deselezione degli articoli da parte degli acquirenti.

  In questo caso, ciascuno degli edge server gestisce ciascun contatore di cassa e il sistema di storage condiviso funge da punto di sincronizzazione centrale.

Servizi finanziari: Sicurezza umana nei chioschi e prevenzione delle frodi

Le organizzazioni bancarie utilizzano l'ai e l'edge computing per innovare e creare esperienze bancarie personalizzate. I chioschi interattivi che utilizzano l'analisi dei dati in tempo reale e l'inferenza ai consentono ora agli ATM non solo di aiutare i clienti a prelevare denaro, ma anche di monitorare in modo proattivo i chioschi attraverso le immagini acquisite dalle telecamere per identificare i rischi per la sicurezza umana o i comportamenti fraudolenti. In questo scenario, i server di calcolo edge e i sistemi storage condivisi sono collegati a chioschi e telecamere interattivi per aiutare le banche a raccogliere ed elaborare i dati con modelli di inferenza ai.

Produzione: Settore 4.0

È iniziata la quarta rivoluzione industriale (Industry 4.0), insieme a trend emergenti come Smart Factory e stampa 3D. Per prepararsi a un futuro basato sui dati, la comunicazione machine-to-machine (M2M) e l'IoT su larga scala sono integrati per una maggiore automazione senza la necessità di un intervento umano. La produzione è già altamente automatizzata e l'aggiunta di funzionalità di ai è una naturale continuazione della tendenza a lungo termine. L'ai consente di automatizzare le operazioni che possono essere automatizzate con l'aiuto di computer Vision e altre funzionalità di ai. È possibile automatizzare il controllo di qualità o le attività che si basano sulla visione umana o sul processo decisionale per eseguire analisi più rapide dei materiali sulle linee di assemblaggio nei piani della fabbrica, in modo da aiutare gli impianti di produzione a soddisfare gli standard ISO richiesti per la gestione della qualità e della sicurezza. In questo caso, ogni edge server di calcolo è connesso a un array di sensori che monitorano il processo di produzione e i modelli di inferenza

aggiornati vengono inviati allo storage condiviso, in base alle necessità.

Telecomunicazioni: Rilevamento della ruggine, ispezione della torre e ottimizzazione della rete

Il settore delle telecomunicazioni utilizza tecniche di visione computerizzata e ai per elaborare immagini che rilevano automaticamente la ruggine e identificano le torri cellulari che contengono corrosione e, di conseguenza, richiedono un'ulteriore ispezione. Negli ultimi anni è aumentato l'utilizzo di immagini drone e modelli ai per identificare regioni distinte di una torre per analizzare ruggine, crepe superficiali e corrosione. La domanda continua a crescere per le tecnologie ai che consentono di ispezionare in modo efficiente l'infrastruttura di telecomunicazione e i ripetitori cellulari, valutarne regolarmente il degrado e ripararli tempestivamente quando necessario.

Inoltre, un altro caso d'utilizzo emergente nel settore delle telecomunicazioni è l'utilizzo di algoritmi ai e ML per prevedere i modelli di traffico dati, rilevare i dispositivi compatibili con il 5G e automatizzare e aumentare la gestione dell'energia MIMO (Multiple-Input and Multiple-Output). L'hardware MIMO viene utilizzato nelle radio tower per aumentare la capacità di rete, ma ciò comporta costi energetici aggiuntivi. I modelli ML per la "modalità di sospensione MIMO" implementati nei siti cellulari possono prevedere l'utilizzo efficiente delle radio e contribuire a ridurre i costi di consumo energetico per gli operatori di reti mobili (MNOS). Le soluzioni di inferenza ai e edge computing aiutano gli MNOS a ridurre la quantità di dati trasmessi avanti e indietro ai data center, ridurre il TCO, ottimizzare le operazioni di rete e migliorare le performance complessive per gli utenti finali.

Panoramica della tecnologia

Questa sezione descrive le basi tecnologiche di questa soluzione ai.

Sistemi NetApp AFF

I sistemi storage NetApp AFF all'avanguardia consentono implementazioni di inferenza ai ai ai edge per soddisfare i requisiti di storage Enterprise con performance leader di settore, flessibilità superiore, integrazione nel cloud e gestione dei dati Best-in-class. Progettati appositamente per la tecnologia flash, i sistemi NetApp AFF aiutano ad accelerare, gestire e proteggere i dati business-critical.

- I sistemi storage entry-level NetApp AFF sono basati su hardware FAS2750 e supporti flash SSD
- Due controller in configurazione ha

I sistemi storage AFF C190 entry-level di NetApp supportano le seguenti funzionalità:

- Un numero massimo di dischi SSD 24x 960 GB
- Due possibili configurazioni:
  - Ethernet (10 GbE): 4 porte 10GBASE-T (RJ-45)
  - Unified (16 GB FC o 10 GbE): 4 porte UTA2 (Unified Target Adapter)
- Capacità effettiva massima di 50,5 TB

> Per i carichi di lavoro NAS, un singolo sistema AFF C190 entry-level supporta un throughput di 4,4 Gbps per letture sequenziali e 230 K IOPS per piccole letture casuali a latenze di 1 ms o inferiori.

NetApp AFF A220

NetApp offre anche altri sistemi storage entry-level che offrono performance e scalabilità superiori per implementazioni su larga scala. Per i carichi di lavoro NAS, un singolo sistema AFF A220 entry-level supporta:

- Throughput di 6,2 Gbps per letture sequenziali
- 375.000 IOPS per piccole letture casuali con latenze di 1 ms o meno
- Numero massimo di dischi SSD 144 x 960 GB, 3,8 TB o 7,6 TB
- AFF A220 è in grado di scalare fino a un massimo di 1 PB di capacità effettiva

NetApp AFF A250

- La capacità effettiva massima è di 35 PB con una scalabilità massima di 2-24 nodi (12 coppie ha)
- Offre un aumento delle performance di ≥ 45% rispetto a AFF A220
- 440.000 letture casuali IOPS @1 ms.
- Basato sull'ultima release di NetApp ONTAP: ONTAP 9.8
- Sfrutta due porte Ethernet da 25 GB per l'interconnessione di ha e cluster

Sistemi NetApp e-Series EF

EF-Series è una famiglia di storage array SAN all-flash entry-level e mid-range in grado di accelerare l'accesso ai dati e di trarne valore più rapidamente con il software NetApp SANtricity. Questi sistemi offrono storage flash SAS e NVMe e offrono IOPS da convenienti a estremi, tempi di risposta inferiori a 100 microsecondi e larghezza di banda fino a 44 Gbps, il che li rende ideali per carichi di lavoro misti e applicazioni esigenti come l'inferenza ai e l'High Performance Computing (HPC).

La figura seguente mostra il sistema storage NetApp EF280.

NetApp EF280

- Supporto FC 32 GB/16 GB, iSCSI 25 GB/10 GB e SAS 12 GB
- La capacità effettiva massima è di 96 dischi per un totale di 1,5 PB
- Throughput di 10 Gbps (letture sequenziali)
- 300.000 IOPS (letture casuali)
- NetApp EF280 è l'array all-flash (AFA) più economico del portfolio NetApp

NetApp EF300

- 24 unità SSD NVMe per una capacità totale di 367 TB
- Opzioni di espansione per un totale di 240x HDD NL-SAS, 96x SSD SAS o una combinazione
- 100 GB di NVMe/IB, NVMe/RoCE, iSER/IB e SRP/IB
- 32 GB NVME/FC, FCP
- 25 GB iSCSI
- 20 Gbps (letture sequenziali)
- 670.000 IOPS (letture casuali)

Per ulteriori informazioni, consultare "Scheda informativa sugli array all-flash NetApp EF-Series EF600, F300, EF570 e EF280".

NetApp ONTAP 9

ONTAP 9.8.1, l'ultima generazione di software per la gestione dello storage NetApp, consente alle aziende di modernizzare l'infrastruttura e passare a un data center cloud-ready. Sfruttando le funzionalità di gestione dei dati leader del settore, ONTAP consente la gestione e la protezione dei dati con un singolo set di strumenti, indipendentemente dalla posizione dei dati. Puoi anche spostare liberamente i dati ovunque siano necessari: Edge, core o cloud. ONTAP 9.8.1 include numerose funzionalità che semplificano la gestione dei dati, accelerano e proteggono i dati critici e abilitano le funzionalità dell'infrastruttura di nuova generazione nelle architetture di cloud ibrido.

Semplifica la gestione dei dati

La gestione dei dati è fondamentale per le operazioni IT aziendali, in modo da utilizzare le risorse appropriate

per applicazioni e set di dati. ONTAP include le seguenti funzionalità per ottimizzare e semplificare le operazioni e ridurre il costo totale delle operazioni:

- **Compattazione dei dati inline e deduplica estesa.** la compattazione dei dati riduce lo spazio sprecato all'interno dei blocchi di storage e la deduplica aumenta significativamente la capacità effettiva. Ciò vale per i dati memorizzati localmente e per i dati a più livelli nel cloud.
- **Qualità del servizio (AQoS) minima, massima e adattiva.** i controlli granulari della qualità del servizio (QoS) aiutano a mantenere i livelli di performance per le applicazioni critiche in ambienti altamente condivisi.
- **NetApp FabricPool.** questa funzione offre il tiering automatico dei dati cold per le opzioni di cloud storage pubblico e privato, tra cui Amazon Web Services (AWS), Azure e la soluzione di storage NetApp StorageGRID. Per ulteriori informazioni su FabricPool, vedere "TR-4598".

Accelera e proteggi i dati

ONTAP 9 offre livelli superiori di performance e protezione dei dati ed estende queste funzionalità nei seguenti modi:

- **Prestazioni e latenza ridotta.** ONTAP offre il throughput più elevato possibile con la latenza più bassa possibile.
- **Protezione dei dati.** ONTAP offre funzionalità di protezione dei dati integrate con gestione comune su tutte le piattaforme.
- **Crittografia dei volumi NetApp (NVE).** ONTAP offre crittografia nativa a livello di volume con supporto per la gestione delle chiavi integrata ed esterna.
- **Multitenancy e autenticazione a più fattori.** ONTAP consente la condivisione delle risorse dell'infrastruttura con i massimi livelli di sicurezza.

Infrastruttura a prova di futuro

ONTAP 9 aiuta a soddisfare le esigenze di business esigenti e in continua evoluzione con le seguenti funzionalità:

- **Scalabilità perfetta e operazioni senza interruzioni.** ONTAP supporta l'aggiunta senza interruzioni di capacità ai controller esistenti e ai cluster scale-out. I clienti possono eseguire l'upgrade alle tecnologie più recenti, come NVMe e 32GB FC, senza costose migrazioni dei dati o interruzioni.
- **Connessione al cloud.** ONTAP è il software di gestione dello storage più connesso al cloud, con opzioni per lo storage software-defined (ONTAP Select) e le istanze native del cloud (NetApp Cloud Volumes Service) in tutti i cloud pubblici.
- **Integrazione con applicazioni emergenti.** ONTAP offre servizi dati di livello Enterprise per piattaforme e applicazioni di prossima generazione, come veicoli autonomi, città intelligenti e Industry 4.0, utilizzando la stessa infrastruttura che supporta le applicazioni aziendali esistenti.

NetApp SANtricity

NetApp SANtricity è progettato per offrire performance, affidabilità e semplicità leader di settore agli array all-flash ibridi e EF-Series. Ottieni il massimo delle performance e dell'utilizzo degli array all-flash ibridi e EF-Series per applicazioni con carichi di lavoro elevati, tra cui analisi dei dati, videosorveglianza e backup e recovery. Con SANtricity, è possibile completare la modifica della configurazione, la manutenzione, l'espansione della capacità e altre attività mentre lo storage rimane online. SANtricity offre inoltre una protezione dei dati superiore, un monitoraggio proattivo e una sicurezza certificata, il tutto accessibile tramite l'interfaccia di System Manager, semplice da utilizzare e integrata. Per ulteriori informazioni, consultare

"Scheda informativa sul software NetApp e-Series SANtricity".

Prestazioni ottimizzate

Il software SANtricity ottimizzato per le performance offre dati, con IOPS elevati, throughput elevato e bassa latenza, a tutte le applicazioni di analisi dei dati, videosorveglianza e backup. Accelera le performance per applicazioni a bassa latenza, IOPS elevati e applicazioni a elevata larghezza di banda e throughput elevato.

Massimizzare l'uptime

Completa tutte le tue attività di gestione mentre lo storage rimane online. Modificare le configurazioni, eseguire la manutenzione o espandere la capacità senza interrompere l'i/O. Ottieni un'affidabilità Best-in-class con funzionalità automatizzate, configurazione online, tecnologia all'avanguardia Dynamic Disk Pools (DPP) e molto altro ancora.

Resto facile

Il software SANtricity offre una protezione dei dati superiore, un monitoraggio proattivo e una sicurezza certificata, il tutto tramite l'interfaccia di System Manager, semplice da utilizzare e integrata. Semplifica le attività di gestione dello storage. Ottieni la flessibilità necessaria per il tuning avanzato di tutti i sistemi storage e-Series. Gestisci il tuo sistema NetApp e-Series, sempre e ovunque. La nostra interfaccia on-box basata sul web ottimizza il tuo workflow di gestione.

Trident di NetApp

"Trident" NetApp è uno storage dinamico open-source orchestrator per Docker e Kubernetes che semplifica la creazione, la gestione e il consumo dello storage persistente. Trident, un'applicazione nativa di Kubernetes, viene eseguita direttamente all'interno di un cluster Kubernetes. Trident consente ai clienti di implementare senza problemi le immagini dei container DL sullo storage NetApp e offre un'esperienza di livello Enterprise per le implementazioni dei container ai. Gli utenti di Kubernetes (come sviluppatori ML e data scientist) possono creare, gestire e automatizzare orchestrazione e cloning per sfruttare le funzionalità avanzate di gestione dei dati di NetApp basate sulla tecnologia NetApp.

Copia e sincronizzazione di NetApp BlueXP

"Copia e sincronizzazione di BlueXP" È un servizio NetApp per una sincronizzazione dei dati rapida e sicura. Sia che tu debba trasferire file tra condivisioni di file SMB o NFS on-premise, NetApp StorageGRID, NetApp ONTAP S3, NetApp Cloud Volumes Service, Azure NetApp Files, Amazon Simple Storage Service (Amazon S3), Amazon Elastic file System (Amazon EFS), BLOB di Azure, Google Cloud Storage, o IBM Cloud Object Storage, BlueXP Copy and Sync sposta i file dove ne hai bisogno in modo rapido e sicuro. Una volta trasferiti, i dati sono completamente disponibili per l'utilizzo sia sull'origine che sulla destinazione. BlueXP Copy e Sync sincronizza costantemente i dati in base alla pianificazione predefinita, spostando solo i delta, in modo da poter ridurre al minimo tempo e denaro necessari per la replica. BlueXP Copy and Sync è un tool software as a service (SaaS) estremamente semplice da configurare e utilizzare. I trasferimenti dei dati attivati da BlueXP Copy e Sync sono effettuati dai broker di dati. Puoi implementare i broker di dati BlueXP Copy e Sync in AWS, Azure, Google Cloud Platform o on-premise.

Server Lenovo ThinkSystem

I server Lenovo ThinkSystem sono dotati di hardware, software e servizi innovativi che risolvono le sfide attuali dei clienti e offrono un approccio di progettazione modulare e evolutivo, adatto allo scopo, per affrontare le sfide del futuro. Questi server si basano su tecnologie Best-in-class e standard di settore, unite a innovazioni Lenovo differenziate per offrire la massima flessibilità possibile nei server x86.

I vantaggi principali dell'implementazione dei server Lenovo ThinkSystem includono:

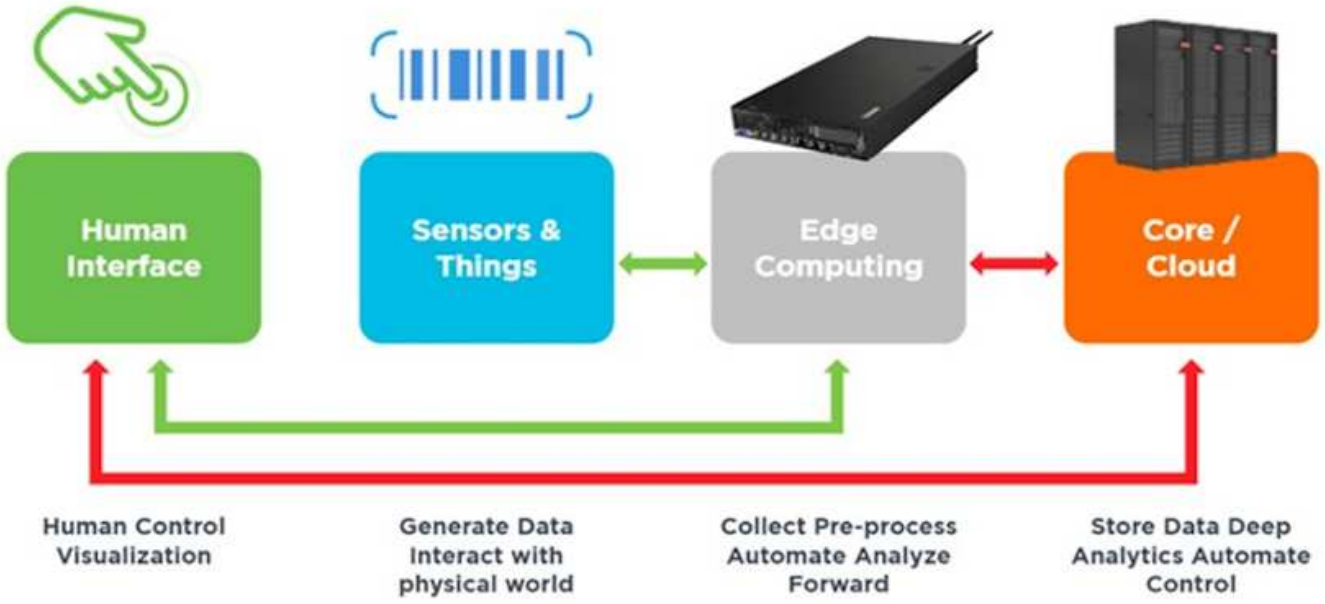
- Design altamente scalabili e modulari per crescere insieme al tuo business
- Resilienza leader del settore per risparmiare ore di costosi downtime non pianificati
- Tecnologie flash veloci per latenze inferiori, tempi di risposta più rapidi e gestione dei dati più intelligente in tempo reale

Nell'area dell'ai, Lenovo sta adottando un approccio pratico per aiutare le aziende a comprendere e adottare i vantaggi di ML e ai per i propri carichi di lavoro. I clienti Lenovo possono esplorare e valutare le offerte Lenovo ai nei Lenovo ai Innovation Center per comprendere appieno il valore del loro caso di utilizzo specifico. Per migliorare il time-to-value, questo approccio incentrato sul cliente offre ai clienti una prova di concetto per le piattaforme di sviluppo di soluzioni pronte all'uso e ottimizzate per l'ai.

Lenovo ThinkSystem SE350 Edge Server

L'edge computing consente di analizzare i dati provenienti dai dispositivi IoT all'edge della rete prima di inviarli al data center o al cloud. Lenovo ThinkSystem SE350, come illustrato nella figura seguente, è progettato per soddisfare i requisiti esclusivi di implementazione alla periferia della rete, con particolare attenzione a flessibilità, connettività, sicurezza e gestibilità remota in un fattore di forma compatto e rinforzato dal punto di vista ambientale.

Dotato del processore Intel Xeon D con la flessibilità di supportare l'accelerazione per i carichi di lavoro ai edge, il SE350 è costruito ad hoc per affrontare la sfida delle implementazioni dei server in una varietà di ambienti esterni al data center.



MLPerf

MLPerf è la suite di benchmark leader del settore per la valutazione delle performance ai. Copre molte aree dell'ai applicata, tra cui classificazione delle immagini, rilevamento degli oggetti, imaging medico e NLP (Natural Language Processing). In questa convalida, abbiamo utilizzato i carichi di lavoro Inference v0.7, che è l'ultima iterazione dell'inferenza MLPerf al completamento di questa convalida. Il "MLPerf Inference v0.7" la suite include quattro nuovi benchmark per data center e sistemi edge:

- **BERT.** rappresentazione del codificatore bidirezionale da Transformers (BERT) ottimizzata per la risposta alle domande utilizzando il set di dati della squadra.
- **DLRM.** Deep Learning Recommendation Model (DLRM) è un modello di personalizzazione e raccomandazione che viene addestrato per ottimizzare i tassi di click-through (CTR).
- **3D U-Net.** l'architettura 3D U-Net viene addestrata sul set di dati Brain Tumor Segmentation (Brats).
- **RNN-T.** il trasduttore di rete neurale ricorrente (RNN-T) è un modello di riconoscimento vocale automatico (ASR) che viene addestrato su un sottoinsieme di LibriSpeech. I risultati e il codice dell'inferenza MLPerf sono pubblicamente disponibili e rilasciati sotto licenza Apache. MLPerf Inference dispone di una divisione Edge, che supporta i seguenti scenari:
- **Single stream.** questo scenario imita i sistemi in cui la reattività è un fattore critico, come le query ai offline eseguite sugli smartphone. Le singole query vengono inviate al sistema e i tempi di risposta vengono registrati. come risultato viene riportata la latenza del 90° percentile di tutte le risposte.
- **Multistream.** questo benchmark è per i sistemi che elaborano input da più sensori. Durante il test, le query vengono inviate a un intervallo di tempo fisso. Viene imposto un vincolo QoS (latenza massima consentita). Il test indica il numero di flussi che il sistema può elaborare rispettando il limite di QoS.
- **Offline.** questo è lo scenario più semplice che copre le applicazioni di elaborazione in batch e la metrica è il throughput in campioni al secondo. Tutti i dati sono disponibili per il sistema e il benchmark misura il tempo necessario per elaborare tutti i campioni.

Lenovo ha pubblicato i punteggi di inferenza MLPerf per SE350 con T4, il server utilizzato in questo documento. Vedere i risultati all'indirizzo "https://mlperf.org/inference-results-0-7/" Nella sezione "Edge, CLOSED Division" della voce 0.7-145.

Piano di test

Questo documento segue l'inferenza MLPerf v0.7 "codice", MLPerf Inference v1.1 "codice", e. "regole". Abbiamo eseguito benchmark MLPerf progettati per l'inferenza ai margini, come definito nella tabella seguente.

Area	Attività	Modello	Dataset	Dimensione QSL	Qualità	Vincolo di latenza multi-stream
Visione	Classificazione delle immagini	Resnet50v1.5	ImageNet (224 x 224)	1024	99% del 32° PQ	50 ms.
Visione	Rilevamento di oggetti (grande)	SSD-ResNet34	COCO (1200 x 1200)	64	99% del 32° PQ	66 ms.
Visione	Rilevamento di oggetti (piccolo)	SSD - MobileNetsv1	COCO (300 x 300)	256	99% del 32° PQ	50 ms.
Visione	Segmentazione delle immagini mediche	UNET 3D	Brat 2019 (224 x 224 x 160)	16	99% e 99.9% del 32° PQ	n/a.
Discorso	Voce-testo	RNNT	Sviluppo di Librispeech-clean	2513	99% del 32° PQ	n/a.
Lingua	Elaborazione della lingua	BERT	Squadra v1.1	10833	99% del 32° PQ	n/a.

Nella tabella seguente sono illustrati gli scenari di benchmark Edge.

Area	Attività	Scenari
Visione	Classificazione delle immagini	Single stream, offline, multistream
Visione	Rilevamento di oggetti (grande)	Single stream, offline, multistream
Visione	Rilevamento di oggetti (piccolo)	Single stream, offline, multistream
Visione	Segmentazione delle immagini mediche	Single stream, offline
Discorso	Voce-testo	Single stream, offline
Lingua	Elaborazione della lingua	Single stream, offline

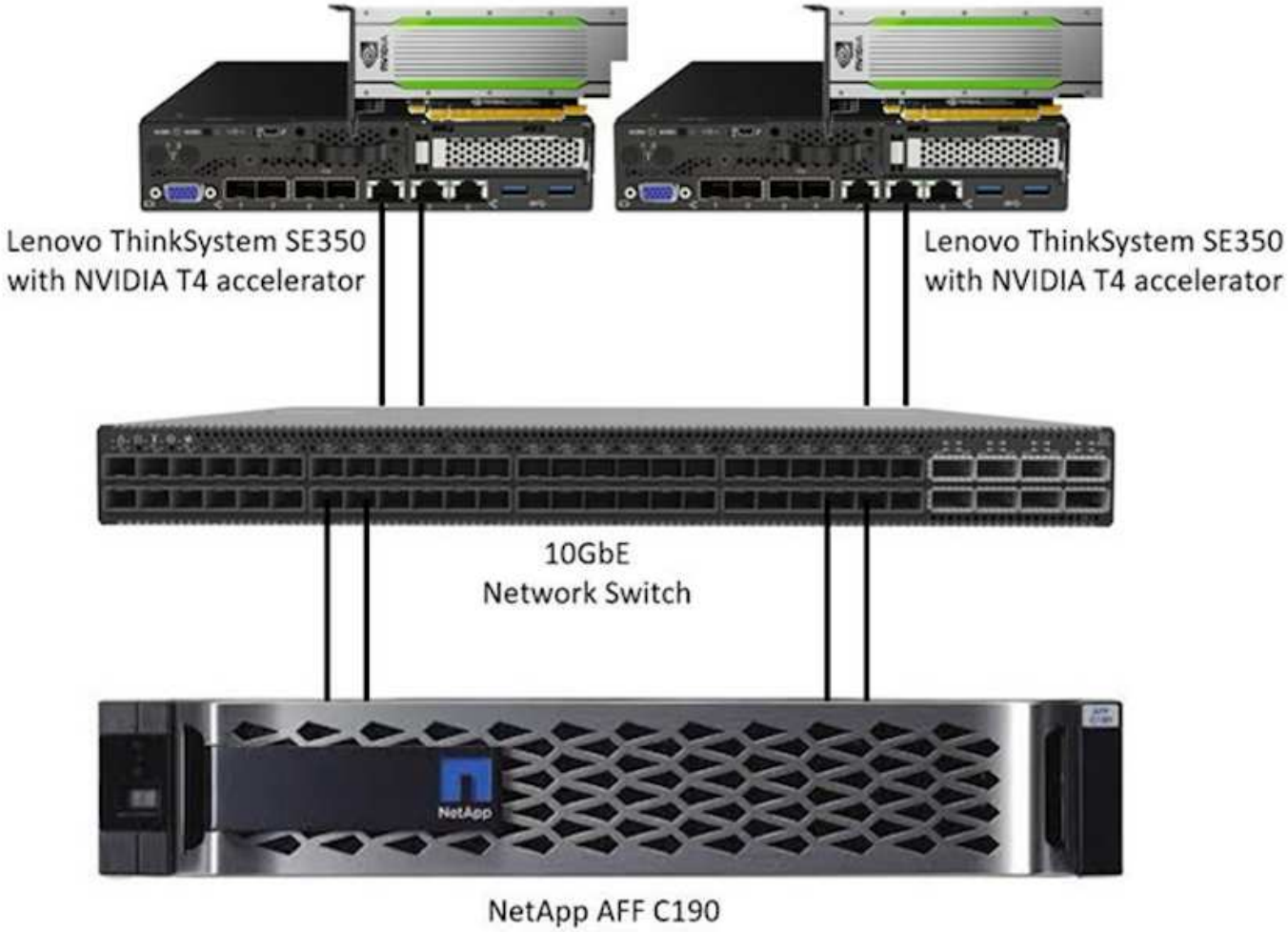
Abbiamo eseguito questi benchmark utilizzando l'architettura di storage di rete sviluppata in questa convalida e confrontato i risultati con quelli delle esecuzioni locali sugli edge server precedentemente inviati a MLPerf. Il confronto serve a determinare l'impatto dello storage condiviso sulle performance di inferenza.

Eseguire il test della configurazione

La seguente figura mostra la configurazione del test. Abbiamo utilizzato il sistema storage NetApp AFF C190 e due server Lenovo ThinkSystem SE350 (ciascuno con un acceleratore NVIDIA T4). Questi componenti sono collegati tramite uno switch di rete 10 GbE. Lo storage di rete contiene set di dati di convalida/test e modelli preformati. I server offrono funzionalità di calcolo e l'accesso allo storage avviene attraverso il protocollo NFS.

Questa sezione descrive le configurazioni testate, l'infrastruttura di rete, il server SE350 e i dettagli relativi al provisioning dello storage. La tabella seguente elenca i componenti di base per l'architettura della soluzione.

Componenti della soluzione	Dettagli
Server Lenovo ThinkSystem	• 2 server SE350 ciascuno con una scheda NVIDIA T4 GPU
	• Ogni server contiene una CPU Intel Xeon D-2123IT con quattro core fisici a 2,20 GHz e 128 GB di RAM
Sistema storage entry-level NetApp AFF (coppia ha)	• Software NetApp ONTAP 9 • 24x 960 GB SSD • Protocollo NFS • Un gruppo di interfacce per controller, con quattro indirizzi IP logici per i punti di montaggio



La seguente tabella elenca la configurazione dello storage: AFF C190 con 2 RU, 24 slot per unità.

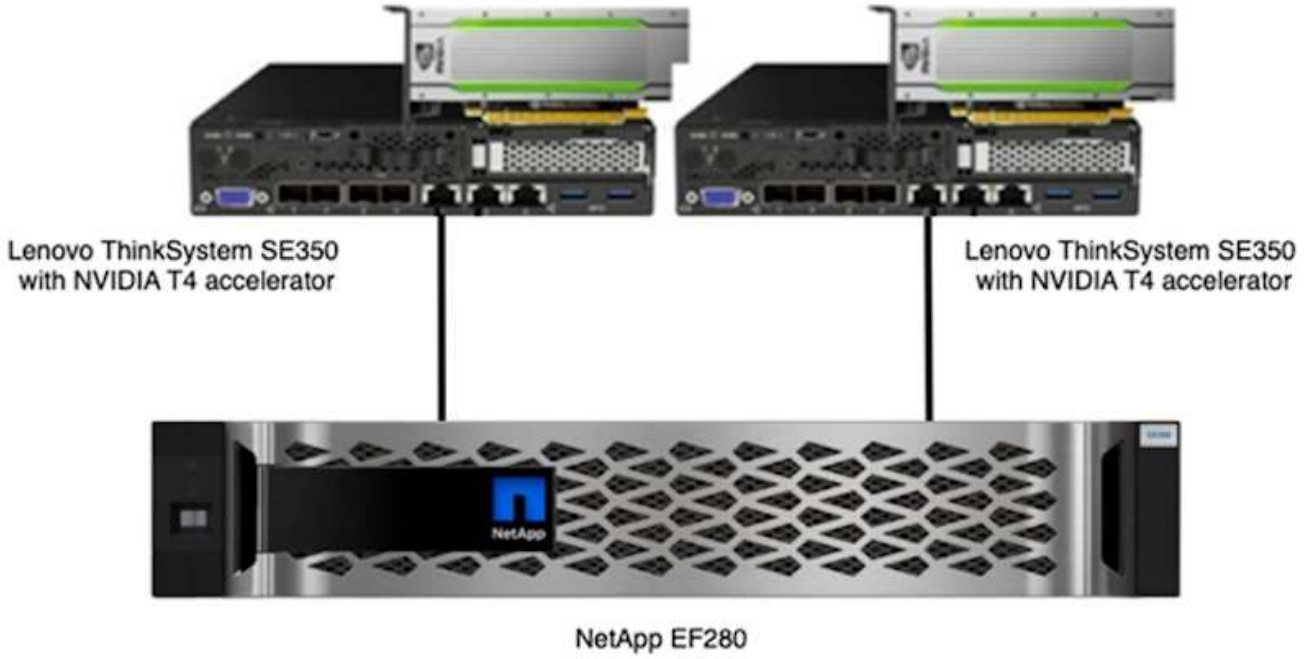
Controller	Aggregato	Volume FlexGroup	Aggregatesize	Volumesize (Volumesize)	Punto di montaggio del sistema operativo
Controller 1	Aggr1	/Netappelenovo_ai_fg	8,42TiB	15 TB	/netapp_lenovo_fg
Controller 2	Aggr2		8,42TiB		

La cartella /netappLenovo_ai_fg contiene i set di dati utilizzati per la convalida del modello.

La figura seguente mostra la configurazione del test. Abbiamo utilizzato il sistema storage NetApp EF280 e due server Lenovo ThinkSystem SE350 (ciascuno con un acceleratore NVIDIA T4). Questi componenti sono collegati tramite uno switch di rete 10 GbE. Lo storage di rete contiene set di dati di convalida/test e modelli preformati. I server offrono funzionalità di calcolo e l'accesso allo storage avviene attraverso il protocollo NFS.

La seguente tabella elenca la configurazione dello storage per EF280.

Controller	Gruppo di volumi	Volume	Volumesize (Volumesize)	DDDP	Metodo di connessione
Controller 1	DDP1	Volume 1	8,42TiB	16 TB	Da SE350-1 a LUN iSCSI 0
Controller 2		Volume 2	8,42TiB		Da SE350-2 a LUN iSCSI 1



Procedura di test

Questa sezione descrive le procedure di test utilizzate per validare questa soluzione.

Configurazione del sistema operativo e dell'inferenza ai

Per AFF C190, abbiamo utilizzato Ubuntu 18.04 con driver NVIDIA e docker con supporto per GPU NVIDIA e abbiamo utilizzato MLPerf "codice" Disponibile come parte dell'invio di Lenovo a MLPerf Inference v0.7.

Per EF280, abbiamo utilizzato Ubuntu 20.04 con driver NVIDIA e docker con supporto per GPU NVIDIA e MLPerf "codice" Disponibile come parte dell'invio di Lenovo a MLPerf Inference v1.1.

Per impostare l'inferenza ai, segui questi passaggi:

1. Scarica i set di dati che richiedono la registrazione, il set di convalida ImageNet 2012, il set di dati Criteo Terabyte e il set di training Brats 2019, quindi decomprimere i file.
2. Creare una directory di lavoro con almeno 1 TB e definire la variabile ambientale `MLPERF_SCRATCH_PATH` facendo riferimento alla directory.

   È necessario condividere questa directory sullo storage condiviso per il caso di utilizzo dello storage di rete o sul disco locale durante il test con dati locali.

3. Esegui il make `prebuild` che crea e avvia il contenitore del docker per le attività di inferenza richieste.

> I seguenti comandi vengono eseguiti tutti dall'interno del contenitore di docker in esecuzione:

- Scarica i modelli ai preformati per le attività di inferenza MLPerf: `make download_model`
- Scarica altri set di dati scaricabili gratuitamente: `make download_data`
- Pre-elaborare i dati: Make `preprocess_data`
- Esecuzione: `make build`.
- Creazione di motori di inferenza ottimizzati per la GPU nei server di calcolo: `make generate_engines`
- Per eseguire i carichi di lavoro di inferenza, eseguire quanto segue (un comando):

```
make run_harness RUN_ARGS="--benchmarks=<BENCHMARKS>
--scenarios=<SCENARIOS>"
```

L'inferenza ai è in esecuzione

Sono stati eseguiti tre tipi di esecuzione:

- Inferenza ai su server singolo utilizzando lo storage locale
- Inferenza ai su server singolo utilizzando lo storage di rete
- Inferenza ai multi-server utilizzando lo storage di rete

Risultati del test

Sono stati eseguiti numerosi test per valutare le performance dell'architettura proposta.

Esistono sei diversi carichi di lavoro (classificazione delle immagini, rilevamento degli oggetti [piccoli], rilevamento degli oggetti [grandi], imaging medico, comunicazione vocale, E Natural Language Processing [NLP]), che è possibile eseguire in tre diversi scenari: Offline, single stream e multistream.

> L'ultimo scenario viene implementato solo per la classificazione delle immagini e il rilevamento degli oggetti.

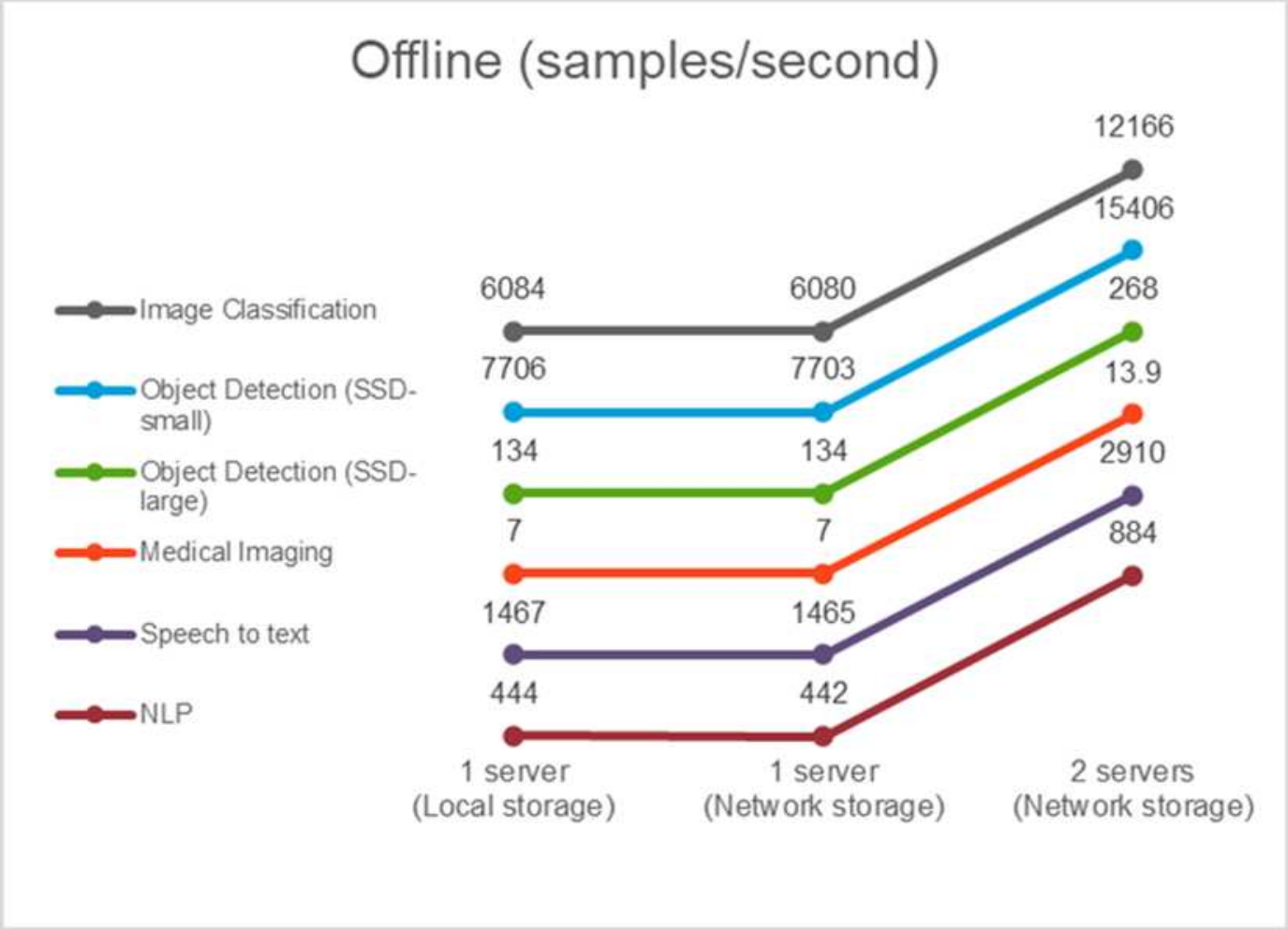
Ciò offre 15 possibili carichi di lavoro, tutti testati in tre diverse configurazioni:

- Server singolo/storage locale
- Storage di rete/server singolo
- Storage di rete/multi-server

I risultati sono descritti nelle sezioni seguenti.

Inferenza ai nello scenario offline per AFF

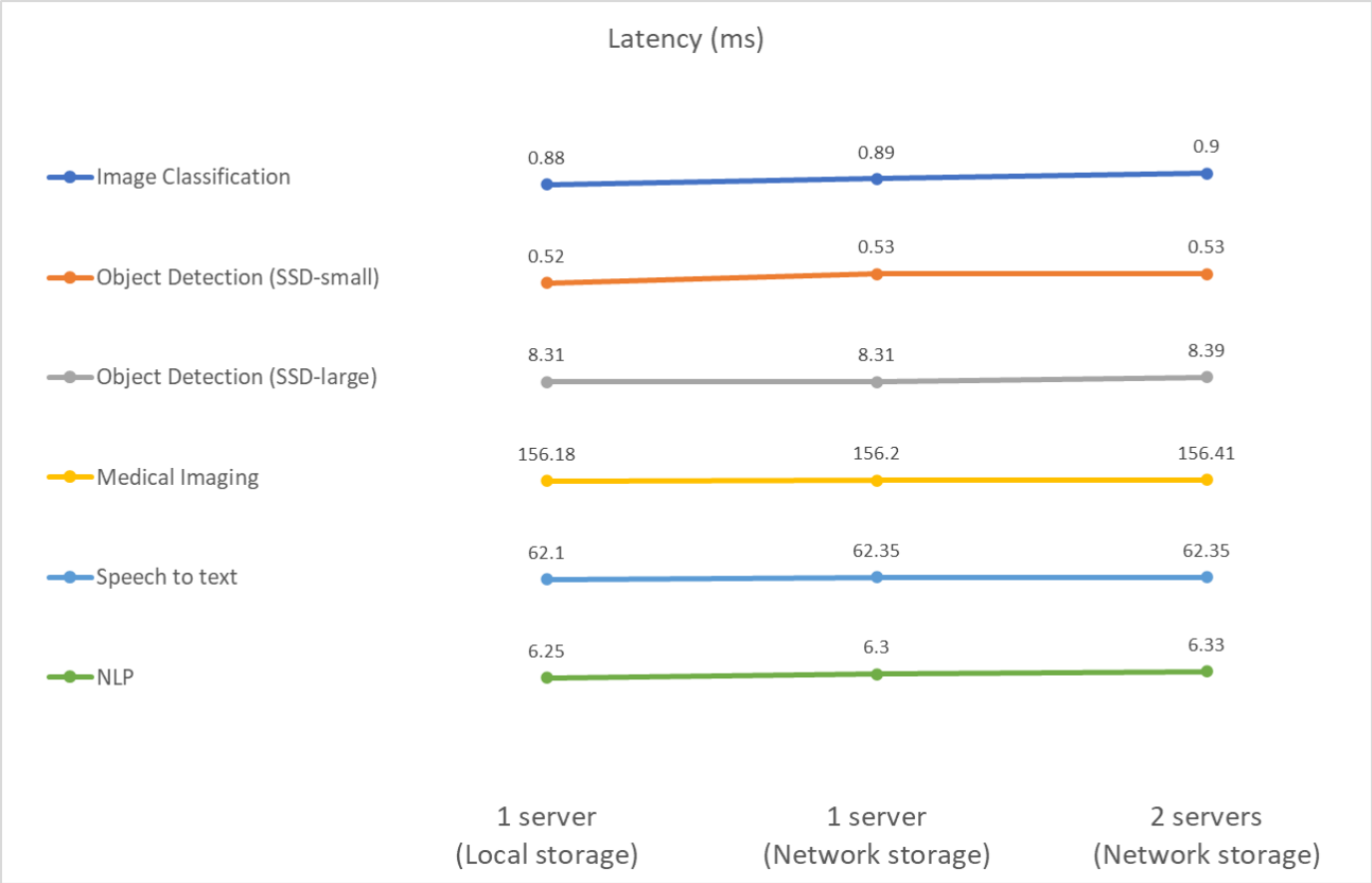
In questo scenario, tutti i dati erano disponibili per il server e il tempo impiegato per elaborare tutti i campioni è stato misurato. I risultati dei test riportano le larghezze di banda in campioni al secondo. Quando sono stati utilizzati più server di calcolo, si riporta la larghezza di banda totale sommata su tutti i server. I risultati per tutti e tre i casi di utilizzo sono mostrati nella figura seguente. Per il caso di due server, segnaliamo la larghezza di banda combinata di entrambi i server.



I risultati mostrano che lo storage di rete non influisce negativamente sulle performance: La modifica è minima e per alcune attività non viene rilevata alcuna. Quando si aggiunge il secondo server, la larghezza di banda totale raddoppia esattamente o, nel peggiore dei casi, la modifica è inferiore all'1%.

Inferenza ai in uno scenario a flusso singolo per AFF

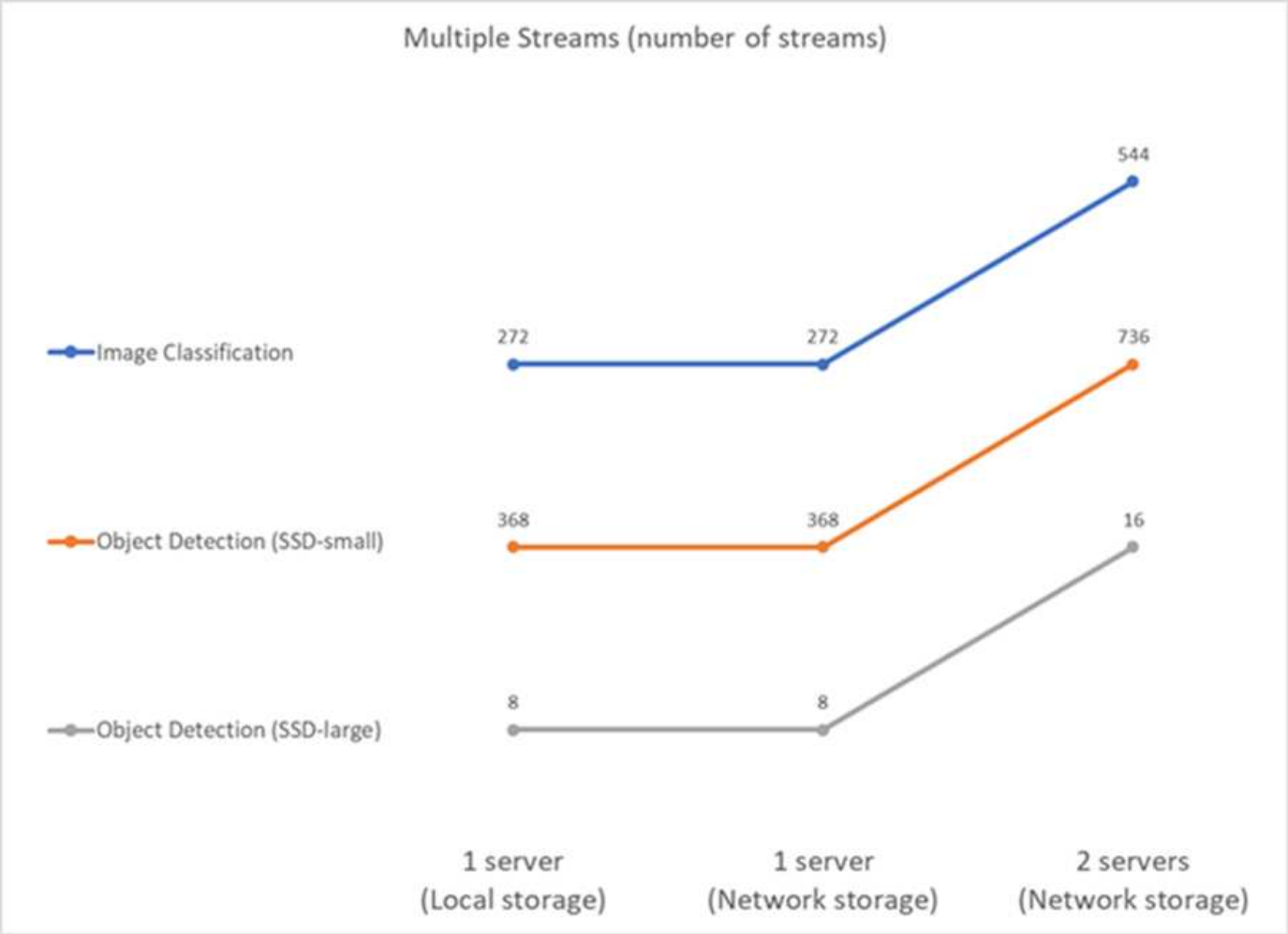
Questo benchmark misura la latenza. Per il caso di più server di calcolo, segnaliamo la latenza media. I risultati per la suite di attività sono riportati nella figura seguente. Per il caso di due server, segnaliamo la latenza media di entrambi i server.



I risultati, ancora una volta, mostrano che lo storage di rete è sufficiente per gestire le attività. La differenza tra storage locale e di rete nel caso di un server è minima o nulla. Allo stesso modo, quando due server utilizzano lo stesso storage, la latenza su entrambi i server rimane la stessa o cambia di molto.

Inferenza ai nello scenario multistream per AFF

In questo caso, il risultato è il numero di flussi che il sistema è in grado di gestire, soddisfacendo al tempo stesso il limite di QoS. Pertanto, il risultato è sempre un numero intero. Per più di un server, viene riportato il numero totale di flussi sommati su tutti i server. Non tutti i carichi di lavoro supportano questo scenario, ma abbiamo eseguito quelli che lo fanno. I risultati dei nostri test sono riassunti nella figura seguente. Per il caso di due server, segnaliamo il numero combinato di flussi da entrambi i server.



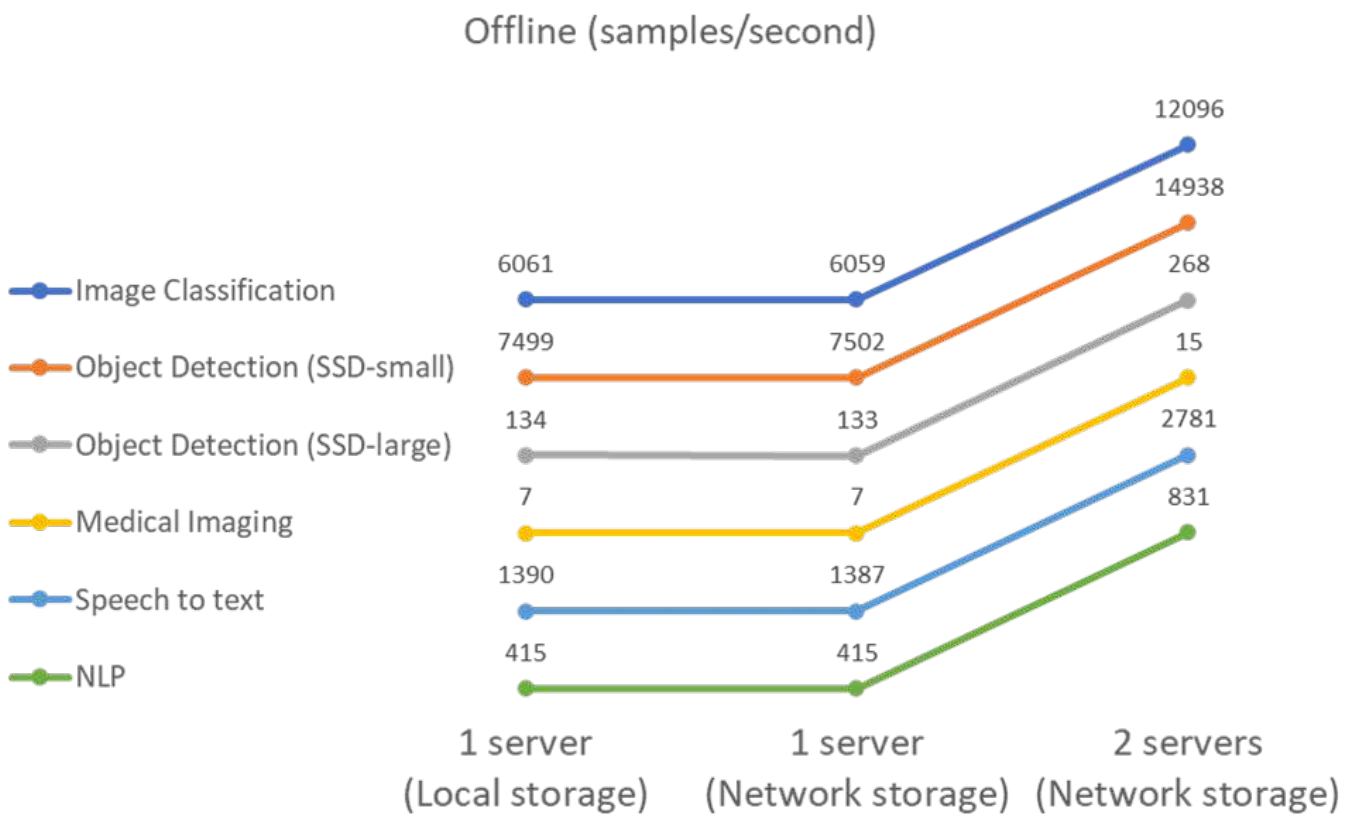
I risultati mostrano le performance perfette dell'installazione: Lo storage locale e di rete offrono gli stessi risultati e l'aggiunta del secondo server raddoppia il numero di flussi gestibili dall'installazione proposta.

Risultati del test per EF

Sono stati eseguiti numerosi test per valutare le performance dell'architettura proposta. Esistono sei diversi carichi di lavoro (classificazione delle immagini, rilevamento degli oggetti [piccoli], rilevamento degli oggetti [grandi], imaging medico, comunicazione vocale, E Natural Language Processing [NLP]), eseguiti in due scenari diversi: Offline e single stream. I risultati sono descritti nelle sezioni seguenti.

Inferenza ai nello scenario offline per EF

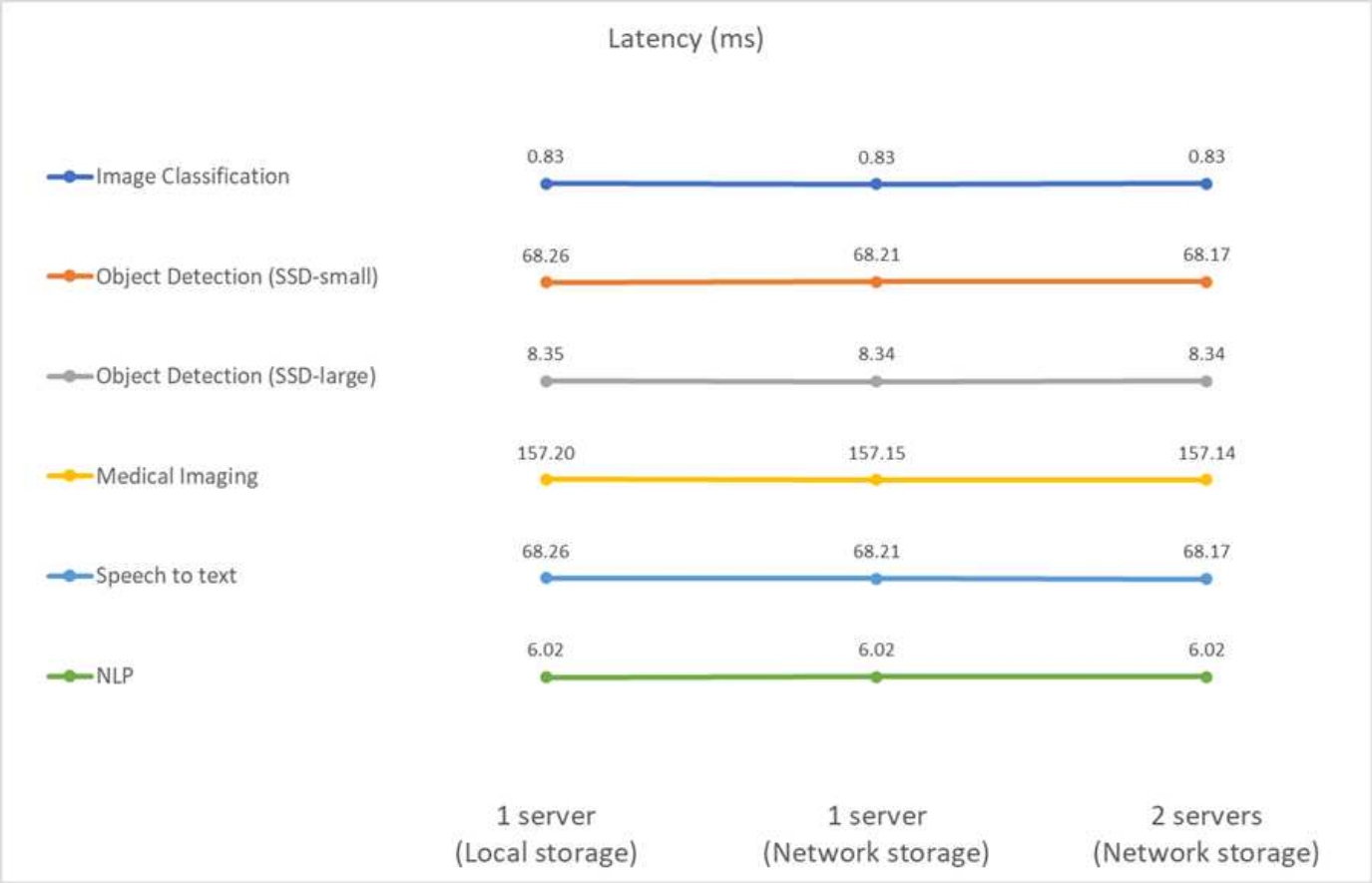
In questo scenario, tutti i dati erano disponibili per il server e il tempo impiegato per elaborare tutti i campioni è stato misurato. I risultati dei test riportano le larghezze di banda in campioni al secondo. Per le esecuzioni a nodo singolo, segnaliamo la media di entrambi i server, mentre per due esecuzioni a server segnaliamo la larghezza di banda totale sommata su tutti i server. I risultati dei casi di utilizzo sono mostrati nella figura seguente.



I risultati mostrano che lo storage di rete non influisce negativamente sulle performance: La modifica è minima e per alcune attività non viene rilevata alcuna. Quando si aggiunge il secondo server, la larghezza di banda totale raddoppia esattamente o, nel peggiore dei casi, la modifica è inferiore all'1%.

Inferenza ai in uno scenario a flusso singolo per EF

Questo benchmark misura la latenza. In tutti i casi, segnaliamo la latenza media su tutti i server coinvolti nelle esecuzioni. Vengono forniti i risultati per la suite di attività.



I risultati mostrano ancora una volta che lo storage di rete è sufficiente per gestire le attività. La differenza tra lo storage locale e di rete nel caso di un server è minima o nulla. Allo stesso modo, quando due server utilizzano lo stesso storage, la latenza su entrambi i server rimane la stessa o cambia di molto.

Opzioni di dimensionamento dell'architettura

È possibile regolare l'impostazione utilizzata per la convalida in modo che si adatti ad altri casi di utilizzo.

Server di calcolo

Abbiamo utilizzato una CPU Intel Xeon D-2123IT, che è il livello più basso di CPU supportato in SE350, con quattro core fisici e TDP da 60 W. Anche se il server non supporta la sostituzione delle CPU, può essere ordinato con una CPU più potente. La CPU più alta supportata è Intel Xeon D-2183IT con 16 core, 100 W a 2,20 GHz. Ciò aumenta notevolmente la capacità di calcolo della CPU. Anche se la CPU non era un collo di bottiglia per l'esecuzione dei carichi di lavoro di inferenza, aiuta nell'elaborazione dei dati e in altre attività correlate all'inferenza. Attualmente, NVIDIA T4 è l'unica GPU disponibile per i casi di utilizzo edge; pertanto, attualmente, non è possibile aggiornare o eseguire il downgrade della GPU.

Storage condiviso

Per il test e la convalida, il sistema NetApp AFF C190, con una capacità di storage massima di 50,5 TB, un throughput di 4,4 Gbps per letture sequenziali e 230 K IOPS per letture casuali di piccole dimensioni, è stato utilizzato per lo scopo di questo documento ed è stato dimostrato adatto per i carichi di lavoro di inferenza edge.

Tuttavia, se si desidera una maggiore capacità di storage o velocità di rete superiori, è consigliabile utilizzare NetApp AFF A220 o. "NetApp AFF A250" sistemi storage. Inoltre, per la convalida della soluzione è stato utilizzato anche il sistema NetApp EF280, con una capacità massima di 1,5 PB e una larghezza di banda di 10 Gbps. Se preferisci una maggiore capacità di storage con una maggiore larghezza di banda, "NetApp EF300" può essere utilizzato.

Conclusione

L'automazione basata sull'ai e l'edge computing sono un approccio leader per aiutare le organizzazioni aziendali a ottenere la trasformazione digitale e massimizzare l'efficienza e la sicurezza delle operazioni. Con l'edge computing, i dati vengono elaborati molto più velocemente perché non devono viaggiare da e verso un data center. Di conseguenza, il costo associato all'invio dei dati avanti e indietro ai data center o al cloud è diminuito. Una latenza inferiore e una velocità maggiore possono essere utili quando le aziende devono prendere decisioni quasi in tempo reale utilizzando modelli di inferenza ai implementati all'edge della rete.

I sistemi storage NetApp offrono performance uguali o migliori rispetto allo storage SSD locale e offrono i seguenti vantaggi a data scientist, data engineer, sviluppatori ai/ML e decision maker aziendali o IT:

- Condivisione semplice dei dati tra sistemi ai, analytics e altri sistemi aziendali critici. Questa condivisione dei dati riduce l'overhead dell'infrastruttura, migliora le performance e ottimizza la gestione dei dati in tutta l'azienda.
- Calcolo e storage scalabili in maniera indipendente per ridurre al minimo i costi e migliorare l'utilizzo delle risorse.
- Workflow di sviluppo e implementazione ottimizzati grazie a copie Snapshot e cloni integrati per spazi di lavoro degli utenti istantanei ed efficienti in termini di spazio, controllo integrato delle versioni e implementazione automatizzata.
- Protezione dei dati di livello Enterprise per disaster recovery e business continuity. La soluzione NetApp e Lenovo presentata in questo documento è un'architettura scalabile e flessibile, ideale per le implementazioni di inferenza ai di livello Enterprise all'edge della rete.

Ringraziamenti

- J.J. Falkanger, Sr Manager, soluzioni HPC e ai, Lenovo
- Dave Arnette, Technical Marketing Engineer, NetApp
- Joey Parnell, Tech Lead e-Series ai Solutions, NetApp
- Cody Harryman, QA Engineer, NetApp

Dove trovare ulteriori informazioni

Per ulteriori informazioni sulle informazioni descritte in questo documento, consultare i seguenti documenti e/o siti Web:

- Pagina del prodotto array NetApp AFF A-Series

  "https://www.netapp.com/data-storage/aff-a-series/"

- Software per la gestione dei dati NetApp ONTAP: Libreria di informazioni ONTAP 9

http://mysupport.netapp.com/documentation/productlibrary/index.html?productID=62286

- TR-4727: Introduzione a NetApp EF-Series

  https://www.netapp.com/pdf.html?item=/media/17179-tr4727pdf.pdf

- Scheda informativa sul software NetApp e-Series SANtricity

  https://www.netapp.com/pdf.html?item=/media/19775-ds-3171-66862.pdf

- Storage persistente NetApp per container - NetApp Trident

  "https://netapp.io/persistent-storage-provisioner-for-kubernetes/"

- MLPerf
  - "https://mlcommons.org/en/"
  - "http://www.image-net.org/"
  - "https://mlcommons.org/en/news/mlperf-inference-v11/"
- Copia e sincronizzazione di NetApp BlueXP

  "https://docs.netapp.com/us-en/occm/concept_cloud_sync.html#how-cloud-sync-works"

- Benchmark TensorFlow

  "https://github.com/tensorflow/benchmarks"

- Lenovo ThinkSystem SE350 Edge Server

  "https://lenovopress.com/lp1168"

- Array di storage flash unificato Lenovo ThinkSystem DM5100F

  "https://lenovopress.com/lp1365-thinksystem-dm5100f-unified-flash-storage-array"

Informazioni sul copyright

Copyright © 2024 NetApp, Inc. Tutti i diritti riservati. Stampato negli Stati Uniti d'America. Nessuna porzione di questo documento soggetta a copyright può essere riprodotta in qualsiasi formato o mezzo (grafico, elettronico o meccanico, inclusi fotocopie, registrazione, nastri o storage in un sistema elettronico) senza previo consenso scritto da parte del detentore del copyright.

Il software derivato dal materiale sottoposto a copyright di NetApp è soggetto alla seguente licenza e dichiarazione di non responsabilità:

IL PRESENTE SOFTWARE VIENE FORNITO DA NETAPP "COSÌ COM'È" E SENZA QUALSIVOGLIA TIPO DI GARANZIA IMPLICITA O ESPRESSA FRA CUI, A TITOLO ESEMPLIFICATIVO E NON ESAUSTIVO, GARANZIE IMPLICITE DI COMMERCIABILITÀ E IDONEITÀ PER UNO SCOPO SPECIFICO, CHE VENGONO DECLINATE DAL PRESENTE DOCUMENTO. NETAPP NON VERRÀ CONSIDERATA RESPONSABILE IN ALCUN CASO PER QUALSIVOGLIA DANNO DIRETTO, INDIRETTO, ACCIDENTALE, SPECIALE, ESEMPLARE E CONSEQUENZIALE (COMPRESI, A TITOLO ESEMPLIFICATIVO E NON ESAUSTIVO, PROCUREMENT O SOSTITUZIONE DI MERCI O SERVIZI, IMPOSSIBILITÀ DI UTILIZZO O PERDITA DI DATI O PROFITTI OPPURE INTERRUZIONE DELL'ATTIVITÀ AZIENDALE) CAUSATO IN QUALSIVOGLIA MODO O IN RELAZIONE A QUALUNQUE TEORIA DI RESPONSABILITÀ, SIA ESSA CONTRATTUALE, RIGOROSA O DOVUTA A INSOLVENZA (COMPRESA LA NEGLIGENZA O ALTRO) INSORTA IN QUALSIASI MODO ATTRAVERSO L'UTILIZZO DEL PRESENTE SOFTWARE ANCHE IN PRESENZA DI UN PREAVVISO CIRCA L'EVENTUALITÀ DI QUESTO TIPO DI DANNI.

NetApp si riserva il diritto di modificare in qualsiasi momento qualunque prodotto descritto nel presente documento senza fornire alcun preavviso. NetApp non si assume alcuna responsabilità circa l'utilizzo dei prodotti o materiali descritti nel presente documento, con l'eccezione di quanto concordato espressamente e per iscritto da NetApp. L'utilizzo o l'acquisto del presente prodotto non comporta il rilascio di una licenza nell'ambito di un qualche diritto di brevetto, marchio commerciale o altro diritto di proprietà intellettuale di NetApp.

Il prodotto descritto in questa guida può essere protetto da uno o più brevetti degli Stati Uniti, esteri o in attesa di approvazione.

LEGENDA PER I DIRITTI SOTTOPOSTI A LIMITAZIONE: l'utilizzo, la duplicazione o la divulgazione da parte degli enti governativi sono soggetti alle limitazioni indicate nel sottoparagrafo (b)(3) della clausola Rights in Technical Data and Computer Software del DFARS 252.227-7013 (FEB 2014) e FAR 52.227-19 (DIC 2007).

I dati contenuti nel presente documento riguardano un articolo commerciale (secondo la definizione data in FAR 2.101) e sono di proprietà di NetApp, Inc. Tutti i dati tecnici e il software NetApp forniti secondo i termini del presente Contratto sono articoli aventi natura commerciale, sviluppati con finanziamenti esclusivamente privati. Il governo statunitense ha una licenza irrevocabile limitata, non esclusiva, non trasferibile, non cedibile, mondiale, per l'utilizzo dei Dati esclusivamente in connessione con e a supporto di un contratto governativo statunitense in base al quale i Dati sono distribuiti. Con la sola esclusione di quanto indicato nel presente documento, i Dati non possono essere utilizzati, divulgati, riprodotti, modificati, visualizzati o mostrati senza la previa approvazione scritta di NetApp, Inc. I diritti di licenza del governo degli Stati Uniti per il Dipartimento della Difesa sono limitati ai diritti identificati nella clausola DFARS 252.227-7015(b) (FEB 2014).

Informazioni sul marchio commerciale

NETAPP, il logo NETAPP e i marchi elencati alla pagina http://www.netapp.com/TM sono marchi di NetApp, Inc. Gli altri nomi di aziende e prodotti potrebbero essere marchi dei rispettivi proprietari.