Inferencia de IA en el perímetro - NetApp con Lenovo ThinkSystem - Diseño de la solución

NetApp Solutions

NetApp
April 24, 2024

This PDF was generated from https://docs.netapp.com/es-es/netapp-solutions/ai/ai-edge-introduction.html on April 24, 2024. Always check docs.netapp.com for the latest.

Tabla de contenidos

Inferencia de IA en el perímetro - NetApp con Lenovo ThinkSystem - Diseño de la solución . . . 1
TR-4886: Inferencia de IA en el Edge - NetApp con Lenovo ThinkSystem - Diseño de la solución . . . 1
Información general de la tecnología . . . 6
Plan de pruebas . . . 13
Configuración de prueba . . . 14
Procedimiento de prueba . . . 16
Resultados de la prueba . . . 17
Opciones de ajuste de tamaño de arquitectura . . . 22
Conclusión . . . 23

Inferencia de IA en el perímetro - NetApp con Lenovo ThinkSystem - Diseño de la solución

TR-4886: Inferencia de IA en el Edge - NetApp con Lenovo ThinkSystem - Diseño de la solución

Sathish Thyagarajan, NetApp Miroslav Hodak, Lenovo

Resumen

Varios supuestos de aplicaciones emergentes, como los sistemas avanzados de asistencia al conductor (ADAS), el sector 4.0, las ciudades inteligentes y el Internet de las cosas (IoT), requieren el procesamiento de flujos de datos continuos con una latencia cercana a cero. Este documento describe una arquitectura de computación y almacenamiento para poner en marcha la inferencia de inteligencia artificial (IA) basada en GPU en controladoras de almacenamiento de NetApp y servidores Lenovo ThinkSystem en un entorno perimetral que satisface estos requisitos. Este documento también proporciona datos de rendimiento para las pruebas de rendimiento de inferencia MLPerf estándares del sector, por lo que evalúa diversas tareas de inferencia en servidores periféricos equipados con GPU T4 de NVIDIA. Investigamos el rendimiento de escenarios de inferencia multisecuencia, sin conexión y con múltiples flujos, y mostramos que la arquitectura con un sistema de almacenamiento en red compartido rentable tiene un alto rendimiento y proporciona un punto central para la gestión de modelos y datos en servidores periféricos múltiples.

Introducción

Las empresas están generando cada vez más volúmenes masivos de datos en el extremo de la red. Con el fin de obtener el máximo valor a partir de sensores inteligentes y datos del Internet de las cosas, las organizaciones buscan una solución de transmisión de eventos en tiempo real que permita el uso de la informática perimetral. Así, pues, cada vez se realizan más tareas informáticas exigentes fuera del perímetro y fuera de los centros de datos. La inferencia de IA es uno de los motores de esta tendencia. Los servidores periféricos proporcionan suficiente potencia computacional para estas cargas de trabajo, sobre todo cuando se utilizan aceleradores, pero el almacenamiento limitado suele ser un problema, especialmente en entornos con varios servidores. En este documento mostramos cómo puede poner en marcha un sistema de almacenamiento compartido en un entorno perimetral y cómo beneficia a las cargas de trabajo de inferencia de IA sin perjudicar el rendimiento.

Este documento describe una arquitectura de referencia para la inferencia de IA en el perímetro. Combina múltiples servidores Lenovo ThinkSystem Edge con un sistema de almacenamiento de NetApp para crear una solución fácil de poner en marcha y gestionar. Está previsto ser una guía de referencia para implementaciones prácticas en diversas situaciones, como la planta de fábrica con múltiples cámaras y sensores industriales, sistemas de punto de venta (POS) en transacciones de venta al por menor o sistemas de autoconducción completa (FSD) que identifican anomalías visuales en vehículos autónomos.

Este documento recoge las pruebas y validación de una configuración de computación y almacenamiento que consta de Lenovo ThinkSystem SE350 Edge Server y un sistema de almacenamiento AFF y EF-Series de NetApp de gama básica. Las arquitecturas de referencia proporcionan una solución eficiente y rentable para puestas en marcha de IA, a la vez que proporcionan servicios de datos completos, protección de datos integrada, escalabilidad fluida y almacenamiento de datos conectado al cloud con ONTAP de NetApp y el software para la gestión de datos SANtricity de NetApp.

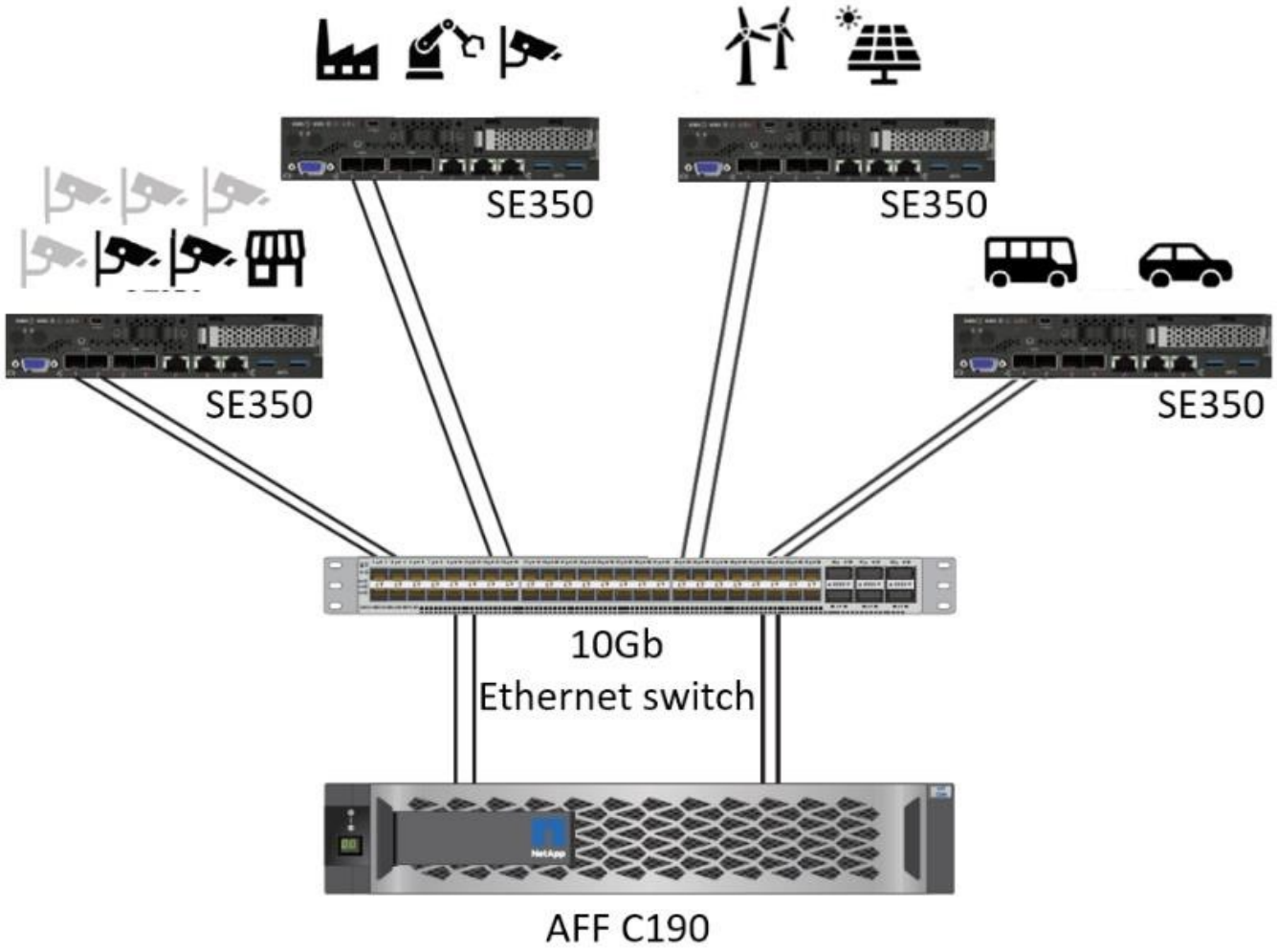
Público objetivo

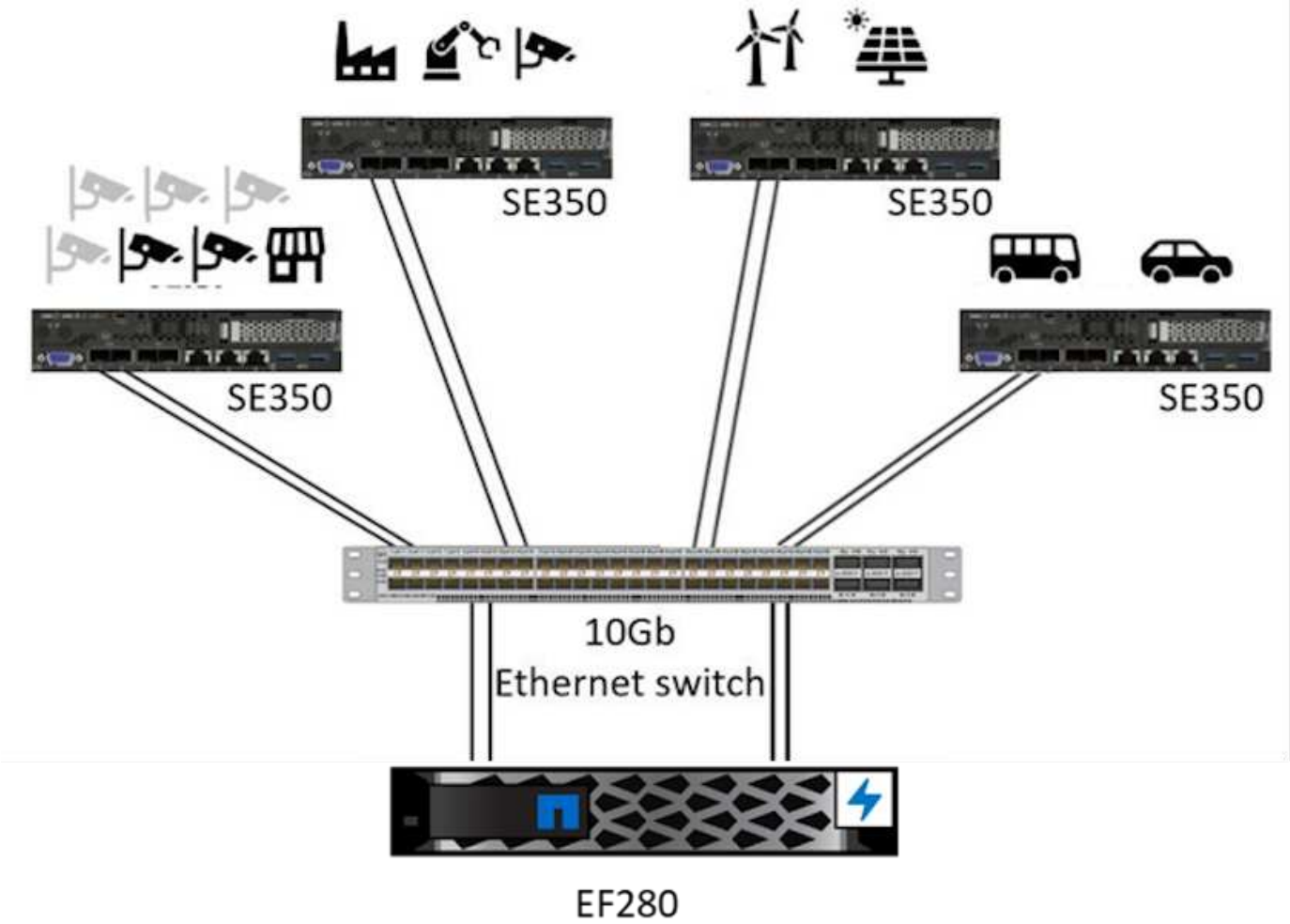
Este documento está dirigido a los siguientes destinatarios:

- Líderes de negocio y arquitectos empresariales que quieren aprovechar la IA en el perímetro.
- Científicos de datos, ingenieros de datos, investigadores DE IA/aprendizaje automático y desarrolladores de sistemas de IA.
- Arquitectos empresariales que diseñan soluciones para el desarrollo de modelos y aplicaciones de IA/ML.
- Científicos e ingenieros de IA que buscan formas eficientes de poner en marcha modelos DE aprendizaje profundo (DL) Y ML.
- Los administradores de dispositivos periféricos y los administradores de servidor perimetral son responsables de la puesta en marcha y la gestión de modelos de inferencia perimetrales.

Arquitectura de la solución

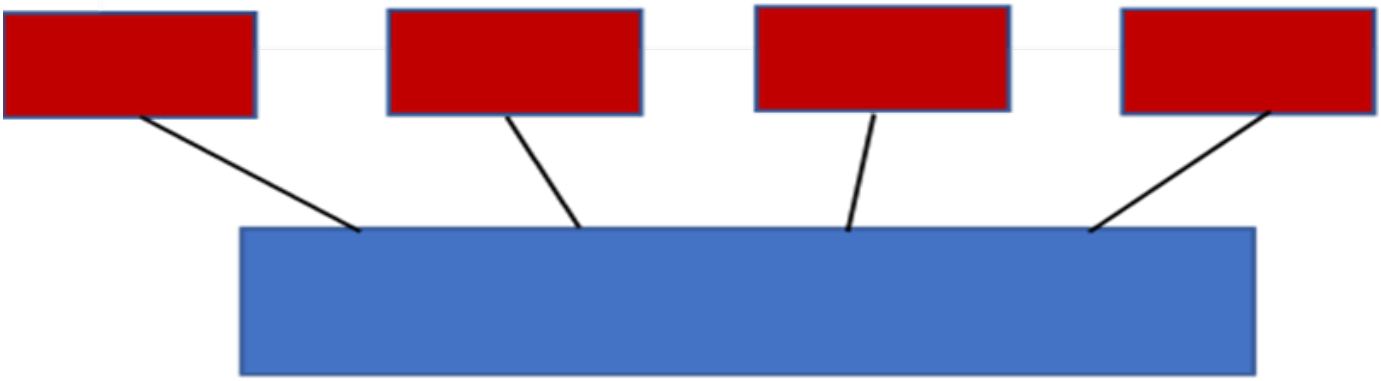
Este servidor Lenovo ThinkSystem y la solución de almacenamiento ONTAP o SANtricity de NetApp están diseñados para gestionar la inferencia de IA en grandes conjuntos de datos mediante la potencia de procesamiento de GPU junto con CPU tradicionales. Esta validación demuestra un alto rendimiento y una gestión de datos óptima con una arquitectura que utiliza uno o varios servidores de borde Lenovo SR350 interconectados con un único sistema de almacenamiento AFF de NetApp, como se muestra en las siguientes dos figuras.





La descripción general de la arquitectura lógica de la siguiente figura muestra las funciones de los elementos de computación y almacenamiento de esta arquitectura. Específicamente, muestra lo siguiente:

- Dispositivos de computación periféricos que realizan inferencia en los datos que recibe de cámaras, sensores, etc.
- Un elemento de almacenamiento compartido que ofrece diversos objetivos:
  - Ofrece una ubicación central para los modelos de inferencia y otros datos que necesitan para realizar la inferencia. Los servidores informáticos acceden al almacenamiento directamente y utilizan los modelos de inferencia en toda la red sin necesidad de copiarlos de forma local.
  - Los modelos actualizados se empujan aquí.
  - Archiva los datos de entrada que los servidores perimetrales reciben para analizarlos posteriormente. Por ejemplo, si los dispositivos periféricos están conectados a las cámaras, el elemento de almacenamiento mantiene los vídeos capturados por las cámaras.



rojo	azul
Sistema de computación Lenovo	Sistema de almacenamiento AFF de NetApp
Dispositivos periféricos que realizan inferencia en las entradas de cámaras, sensores, etc.	Almacenamiento compartido que mantiene los modelos de inferencia y los datos de los dispositivos periféricos para su posterior análisis.

Esta solución de NetApp y Lenovo ofrece las siguientes ventajas clave:

- GPU acelera la computación en el extremo.
- Implementación de varios servidores periféricos, respaldados y gestionado desde un almacenamiento compartido.
- Protección de datos sólida para cumplir con los objetivos de punto de recuperación (RPO) y los objetivos de tiempo de recuperación (RTO) bajos sin pérdida de datos.
- Gestión de datos optimizada con copias Snapshot y clones de NetApp para optimizar los flujos de trabajo de desarrollo.

Cómo utilizar esta arquitectura

Este documento valida el diseño y el rendimiento de la arquitectura propuesta. Sin embargo, no hemos probado ciertas piezas a nivel de software, como la gestión y sincronización de datos de contenedores, cargas de trabajo o modelos con cloud o centros de datos en las instalaciones, ya que son específicas de un escenario de puesta en marcha. Aquí existen varias opciones.

En el nivel de gestión de contenedores, la gestión de contenedores de Kubernetes es una buena opción y es compatible con una versión totalmente ascendente (Canonical) o con una versión modificada adecuada para implementaciones empresariales (Red Hat). La "Plano de control de IA de NetApp" Que utiliza Trident de NetApp y los recientemente añadidos "Kit de herramientas de operaciones de datos de NetApp" Proporciona trazabilidad, funciones de gestión de datos, interfaces y herramientas integradas para que los científicos e ingenieros de datos se integren con el sistema de almacenamiento de NetApp. Kubeflow, el kit DE herramientas ML para Kubernetes, proporciona funcionalidades de IA adicionales junto con compatibilidad para el control de versiones de modelos y KFServing en varias plataformas como TensorFlow Serving o NVIDIA Triton inferpensado Server. Otra opción es la plataforma NVIDIA EGX, que proporciona gestión de las cargas de trabajo junto con acceso a un catálogo de contenedores de inferencia de IA habilitados para GPU. Sin embargo, es posible que estas opciones requieran gran esfuerzo y experiencia para ponerlas en producción y es posible que requieran la ayuda de un proveedor de software independiente (ISV) o un asesor de terceros.

Áreas de soluciones

La ventaja clave de la inferencia de IA y la computación perimetral es la capacidad de que los dispositivos calculen, procesen y analicen datos con un alto nivel de calidad sin latencia. Hay demasiados ejemplos de casos de uso de computación perimetral que describir en este documento, pero aquí hay algunos ejemplos destacados:

Automóviles: Vehículos autónomos

La ilustración clásica de la informática avanzada se encuentra en los sistemas avanzados de asistencia al conductor (ADAS) en vehículos autónomos (AV). La IA en vehículos sin conductor debe procesar rápidamente una gran cantidad de datos procedentes de cámaras y sensores para garantizar su seguridad. Tomar demasiado tiempo para interpretar entre un objeto y un humano puede significar la vida o la muerte, por lo tanto poder procesar los datos lo más cerca posible del vehículo es crucial. En este caso, uno o varios servidores de computación periféricos se encarga de las entradas de cámaras, RADAR, LiDAR y otros sensores, mientras que el almacenamiento compartido contiene modelos de inferencia y almacena datos de entrada de los sensores.

Atención sanitaria: Monitorización de pacientes

Uno de los mayores impactos de la IA y la informática perimetral es su capacidad para mejorar la supervisión continua de pacientes para enfermedades crónicas, tanto en las unidades de cuidados intensivos como en las unidades de cuidados intensivos (UCI). Los datos de los dispositivos periféricos que supervisan los niveles de insulina, la respiración, la actividad neurológica, el ritmo cardíaco y las funciones gastrointestinales requieren un análisis instantáneo de los datos que deben ser objeto de acciones inmediatas porque hay poco tiempo para actuar y salvar la vida de alguien.

Venta al por menor: Pago sin cajero

La computación perimetral puede ayudar a los minoristas a reducir el tiempo de salida y aumentar el tráfico de pies. Los sistemas sin cajero admiten varios componentes, como los siguientes:

- Autenticación y acceso. Conectar el comprador físico a una cuenta validada y permitir el acceso al espacio de venta al por menor.
- Supervisión de inventario. Utilizar sensores, etiquetas RFID y sistemas de visión computarizada para confirmar la selección o deselección de artículos por parte de los compradores.

  Aquí, cada uno de los servidores perimetrales gestiona cada contador de retirada y el sistema de almacenamiento compartido sirve como punto de sincronización central.

Servicios financieros: Seguridad humana en quioscos y prevención del fraude

Las organizaciones bancarias utilizan la IA y la informática perimetral para innovar y crear experiencias bancarias personalizadas. Los quioscos interactivos, mediante el análisis de datos en tiempo real y la inferencia de IA, permiten ahora a los cajeros automáticos no sólo ayudar a los clientes a retirar el dinero, sino también supervisar de forma proactiva los quioscos a través de las imágenes capturadas con las cámaras para identificar el riesgo para la seguridad humana o el comportamiento fraudulento. En este escenario, los servidores periféricos informáticos y los sistemas de almacenamiento compartido se conectan a quioscos y cámaras interactivos para ayudar a los bancos a recopilar y procesar datos con modelos de inferencia de IA.

Fabricación: Industria 4.0

La cuarta revolución industrial (Industry 4.0) ha comenzado, junto con tendencias emergentes como Smart Factory e impresión 3D. Con el fin de prepararse para un futuro impulsado por los datos, la comunicación y el

Internet de las cosas (M2M) a gran escala están integrados para una mayor automatización sin necesidad de intervención humana. La fabricación ya está altamente automatizada y añadir características de IA es una continuación natural de la tendencia a largo plazo. La IA permite automatizar operaciones que se pueden automatizar con la ayuda de la visión computarizada y otras funcionalidades de IA. Puede automatizar el control de calidad o las tareas que se basan en la visión humana o en la toma de decisiones para realizar análisis más rápidos de materiales en líneas de ensamblaje en plantas de fabricación para ayudar a las plantas de fabricación a cumplir con los estándares ISO requeridos de gestión de la seguridad y la calidad. Aquí, cada servidor perimetral informático está conectado a una matriz de sensores que supervisan el proceso de fabricación, y cuando es necesario, los modelos de inferencia actualizados se ven empujados al almacenamiento compartido.

Telecomunicaciones: Detección de óxido, inspección de torre y optimización de la red

La industria de las telecomunicaciones utiliza técnicas de IA y visión informática para procesar imágenes que detectan automáticamente el óxido e identifican las torres celulares que contienen corrosión y, por lo tanto, requieren una inspección adicional. El uso de imágenes con drones y modelos de IA para identificar regiones distintas de una torre para analizar la oxidación, las grietas superficiales y la corrosión ha aumentado en los últimos años. La demanda continúa creciendo para tecnologías de IA que permiten inspeccionar eficientemente la infraestructura de telecomunicaciones y las torres de células, evaluar periódicamente la degradación y reparar rápidamente cuando sea necesario.

Además, otro caso de uso emergente en las telecomunicaciones es el uso de algoritmos de IA y ML para predecir patrones de tráfico de datos, detectar dispositivos compatibles con 5G y automatizar y aumentar la gestión energética de múltiples entradas y salidas múltiples (MIMO). El hardware de MIMO se utiliza en las torres de radio para aumentar la capacidad de la red; sin embargo, esto viene con costos de energía adicionales. Los modelos ML para “modo de suspensión MIMO” implementados en las zonas de células pueden predecir el uso eficiente de las radios y ayudar a reducir los costes de consumo de energía para los operadores de redes móviles (MNO). Las soluciones de computación avanzada y de inferencia de IA ayudan a las MNO a reducir la cantidad de datos transmitidos hacia los centros de datos, reducir su TCO, optimizar las operaciones de red y mejorar el rendimiento general de los usuarios finales.

Información general de la tecnología

Esta sección describe la base tecnológica de esta solución de IA.

Sistemas AFF de NetApp

Los sistemas de almacenamiento AFF de NetApp más innovadores permiten que las puestas en marcha de inferencia de IA en el perímetro satisfagan los requisitos de almacenamiento de la empresa con un rendimiento líder del sector, una flexibilidad superior, integración con el cloud y la mejor gestión de datos de su clase. Los sistemas AFF de NetApp han sido diseñados específicamente para flash y ayudan a acelerar, gestionar y proteger los datos esenciales para la empresa.

- Los sistemas de almacenamiento AFF de NetApp de gama básica se basan en hardware FAS2750 y medios flash SSD
- Dos controladoras en configuración de alta disponibilidad

Los sistemas de almacenamiento C190 de gama básica de AFF admiten las siguientes funciones:

- Un número máximo de unidades de estado sólido de 960 GB
- Dos configuraciones posibles:
  - Ethernet (10 GbE): 4 puertos 10GBASE-T (RJ-45)
  - Unificado (FC de 16 GB o 10 GbE): 4 puertos de adaptador de destino unificado 2 (UTA2)
- Una capacidad efectiva máxima de 50,5 TB

> En el caso de cargas de trabajo NAS, un único sistema AFF C190 de gama básica admite un rendimiento de 4,4 GB/s para lecturas secuenciales y 230 000 IOPS para lecturas aleatorias pequeñas con latencias de 1 ms o menos.

AFF A220 de NetApp

NetApp también ofrece otros sistemas de almacenamiento de gama básica que proporcionan un mayor rendimiento y escalabilidad para las puestas en marcha a gran escala. En el caso de cargas de trabajo NAS, un único sistema AFF A220 de gama básica admite:

- Rendimiento de 6,2 Gbps para lecturas secuenciales
- 3375 000 IOPS para lecturas aleatorias pequeñas con latencias de 1 ms o menos
- Recuento máximo de unidades de SSD 144 TB, 3,8 TB o 7,6 TB
- AFF A220 es escalable a más de 1 PB de capacidad efectiva

AFF A250 de NetApp

- La capacidad efectiva máxima es de 35 PB con el escalado horizontal máximo de 2-24 nodos (12 parejas de alta disponibilidad).
- Ofrece un aumento del rendimiento de ≥ 45 % con respecto a AFF A220
- 440 000 IOPS lecturas aleatorias a 1 ms
- Basado en el último lanzamiento de ONTAP de NetApp: ONTAP 9.8

- Aprovecha dos Ethernet de 25 GB para alta disponibilidad e interconexión de clúster

Sistemas E-Series EF de NetApp

EF-Series es una familia de cabinas DE almacenamiento SAN all-flash de gama básica y media que pueden acelerar el acceso a sus datos y ayudarle a obtener valor de ella más rápido con el software SANtricity de NetApp. Estos sistemas ofrecen almacenamiento flash SAS y NVMe y le proporcionan IOPS asequibles hasta extremas, tiempos de respuesta inferiores a 100 microsegundos y un ancho de banda de hasta 44 Gbps, lo cual los convierte en ideales para cargas de trabajo mixtas y aplicaciones exigentes como la inferencia de IA y la computación de alto rendimiento (HPC).

La siguiente figura muestra el sistema de almacenamiento EF280 de NetApp.

EF280 de NetApp

- Compatibilidad con FC de 32 GB/16 GB, iSCSI de 25 GB/10 GB y SAS de 12 GB
- La capacidad efectiva máxima es 96 unidades que totalizan 1,5 PB
- Rendimiento de 10 Gbps (lecturas secuenciales)
- 300K IOPS (lectura aleatoria)
- El sistema EF280 de NetApp es la cabina all-flash (AFA) más económica de la cartera de NetApp

EF300 de NetApp

- 24 unidades SSD NVMe para una capacidad total de 367 TB
- Opciones de expansión con un total de 240 HDD NL-SAS, 92 SSD SAS o una combinación de ambos
- 100 GB NVMe/IB, NVMe/roce, Iser/IB y SRP/IB
- NVME DE 32 GB/FC, FCP
- ISCSI de 25 GB
- 20 Gbps (lecturas secuenciales)
- 670K IOPS (lecturas aleatorias)

Para obtener más información, consulte "Especificaciones técnicas de las cabinas all-flash EF-Series de NetApp EF600, F300, EF570 y EF280".

ONTAP 9 de NetApp

ONTAP 9.8.1, la última generación del software de gestión del almacenamiento de NetApp, permite a las empresas modernizar su infraestructura y realizar la transición a un centro de datos preparado para el cloud. ONTAP ofrece las mejores capacidades de gestión de datos y permite la gestión y protección de los datos con un solo conjunto de herramientas, sin importar dónde residan. También puede mover los datos libremente a donde sea necesario: El perímetro, el núcleo o el cloud. ONTAP 9.8.1 incluye numerosas funciones que simplifican la gestión de datos, aceleran y protegen los datos esenciales y permiten disfrutar de funcionalidades de infraestructura de nueva generación en arquitecturas de cloud híbrido.

Simplificar la gestión de los datos

La gestión de los datos es crucial para las operaciones TECNOLÓGICAS empresariales, de modo que se utilicen recursos apropiados para las aplicaciones y conjuntos de datos. ONTAP incluye las siguientes funciones para facilitar y simplificar las operaciones, y reducir el coste total de las operaciones:

- **Compactación de datos inline y deduplicación expandida.** la compactación de datos reduce el espacio perdido dentro de los bloques de almacenamiento, y la deduplicación aumenta significativamente la capacidad efectiva. Esto es aplicable a los datos almacenados localmente y a los datos organizados en niveles en el cloud.
- **Calidad de servicio (AQoS) mínima, máxima y adaptativa.** los controles de calidad de servicio (QoS) granulares ayudan a mantener los niveles de rendimiento para aplicaciones críticas en entornos altamente compartidos.
- **FabricPool de NetApp.** esta función ofrece una organización automática por niveles de datos inactivos a opciones de almacenamiento en cloud privado o público, como la solución de almacenamiento Amazon Web Services (AWS), Azure y StorageGRID de NetApp. Para obtener más información sobre FabricPool, consulte "TR-4598".

Acelere y proteja sus datos

ONTAP 9 ofrece niveles superiores de rendimiento y protección de datos, y amplía estas capacidades de las siguientes maneras:

- **Rendimiento y menor latencia.** ONTAP ofrece el rendimiento más alto posible con la menor latencia posible.
- **Protección de datos.** ONTAP ofrece capacidades integradas de protección de datos con administración común en todas las plataformas.
- **Cifrado de volumen de NetApp (NVE).** ONTAP ofrece cifrado nativo a nivel de volumen con compatibilidad para gestión de claves incorporada y externa.
- **Multitenancy y autenticación multifactor.** ONTAP permite compartir recursos de infraestructura con los niveles más altos de seguridad.

Infraestructura preparada para futuros retos

ONTAP 9 ayuda a satisfacer las exigentes y siempre cambiantes necesidades de la empresa con las siguientes funciones:

- **Escalado sencillo y operaciones no disruptivas.** ONTAP admite la adición sin interrupciones de capacidad a las controladoras existentes y a los clústeres de escalado horizontal. Los clientes pueden empezar a utilizar tecnologías punteras como NVMe y FC 32 GB, sin necesidad de realizar costosas migraciones de datos y sin cortes.
- **Conexión en cloud.** ONTAP es el software de gestión del almacenamiento con mejor conexión en cloud,

con opciones de almacenamiento definido por software (ONTAP Select) e instancias nativas del cloud (Cloud Volumes Service de NetApp) en todos los clouds públicos.

- **Integración con aplicaciones emergentes.** ONTAP ofrece servicios de datos de clase empresarial para plataformas y aplicaciones de última generación, como vehículos autónomos, ciudades inteligentes e Industria 4.0, utilizando la misma infraestructura que soporta las aplicaciones empresariales existentes.

SANtricity de NetApp

SANtricity de NetApp está diseñado para ofrecer un rendimiento líder del sector, fiabilidad y simplicidad en las cabinas all-flash EF-Series y flash híbrido. Logre el máximo rendimiento y aprovechamiento de sus cabinas all-flash EF-Series y flash híbridas de E-Series para aplicaciones de cargas de trabajo pesadas, como análisis de datos, videovigilancia y backup y recuperación de datos. Con SANtricity, los ajustes de configuración, el mantenimiento, la expansión de la capacidad y otras tareas se pueden realizar mientras el almacenamiento sigue online. SANtricity también ofrece una protección de datos superior, supervisión proactiva y seguridad certificada: Todo accesible a través de la sencilla interfaz integrada de System Manager. Para obtener más información, consulte "Especificaciones técnicas del software SANtricity para E-Series de NetApp".

Rendimiento optimizado

El software SANtricity optimizado para el rendimiento ofrece datos —con una alta tasa de IOPS, un alto rendimiento y una baja latencia— a todas sus aplicaciones de análisis de datos, videovigilancia y backup. Acelere el rendimiento en aplicaciones con baja latencia y una alta tasa de IOPS, y en aplicaciones con requisitos de amplio ancho de banda y un rendimiento alto.

Maximice el tiempo de actividad

Complete todas las tareas de gestión mientras el almacenamiento sigue en línea. Modifique las configuraciones, realice tareas de mantenimiento o amplíe la capacidad sin interrumpir el flujo de I/O. Proporcione los mayores niveles de fiabilidad posibles con funciones automatizadas, opciones de configuración en línea, la tecnología de pools de discos dinámicos (DPP) de vanguardia, entre otras.

Esté tranquilo

El software SANtricity ofrece una protección de datos superior, supervisión proactiva y seguridad certificada: Todo a través de la sencilla interfaz incluida de System Manager. Simplifique las tareas de gestión del almacenamiento. Obtenga la flexibilidad que necesita para realizar un mejor ajuste de todos los sistemas de almacenamiento E-Series. Gestione su sistema E-Series de NetApp en todo momento y en cualquier lugar. Nuestra interfaz integrada basada en web optimiza el flujo de trabajo de gestión.

Trident de NetApp

"Trident" De NetApp es un orquestador de almacenamiento dinámico de código abierto para Docker y Kubernetes que simplifica la creación, la gestión y el consumo de almacenamiento persistente. Trident, una aplicación nativa de Kubernetes, se ejecuta directamente dentro de un clúster de Kubernetes. Trident permite que los clientes implementen sin problemas imágenes de contenedores de DL en el almacenamiento de NetApp y proporciona una experiencia de clase empresarial para implementaciones de contenedores de IA. Los usuarios de Kubernetes (como desarrolladores DE ML y científicos de datos) pueden crear, gestionar y automatizar la orquestación y el clonado para aprovechar las funcionalidades de gestión de datos avanzadas de NetApp, impulsadas por la tecnología de NetApp.

Copia y sincronización de NetApp BlueXP

"Copia y sincronización de BlueXP" Es un servicio de NetApp que ofrece una sincronización de datos rápida y

segura. Ya tenga que transferir archivos entre recursos compartidos de archivos NFS o SMB en las instalaciones, NetApp StorageGRID, NetApp ONTAP S3, NetApp Cloud Volumes Service, Azure NetApp Files, Amazon Simple Storage Service (Amazon S3), Amazon Elastic File System (Amazon EFS), Azure Blob, Google Cloud Storage, o IBM Cloud Object Storage, BlueXP Copy and Sync mueve los archivos a donde los necesites de forma rápida y segura. Una vez transferidos los datos, estarán completamente disponibles para su uso tanto en origen como en destino. BlueXP Copy and Sync sincroniza los datos de forma continua en función de tu programación predefinida, moviendo solo los deltas, por lo que se reducen al mínimo el tiempo y el dinero que se invierten en la replicación de datos. Copia y sincronización de BlueXP es una herramienta de software como servicio (SaaS) extremadamente sencilla de configurar y usar. Las transferencias de datos activadas por BlueXP Copy and Sync se llevan a cabo por agentes de datos. Puedes poner en marcha agentes de datos de BlueXP Copy y Sync en AWS, Azure, Google Cloud Platform o en las instalaciones.

Servidores Lenovo ThinkSystem

Los servidores Lenovo ThinkSystem incluyen hardware, software y servicios innovadores que resuelven los desafíos actuales de los clientes y ofrecen un enfoque de diseño modular, evolutivo y adecuado para su propósito para afrontar los desafíos del futuro. Estos servidores se capitalizan en las mejores tecnologías estándar del sector, junto con innovaciones diferenciadas de Lenovo, para proporcionar la mayor flexibilidad posible en servidores x86.

Las ventajas clave de la implementación de servidores Lenovo ThinkSystem incluyen:

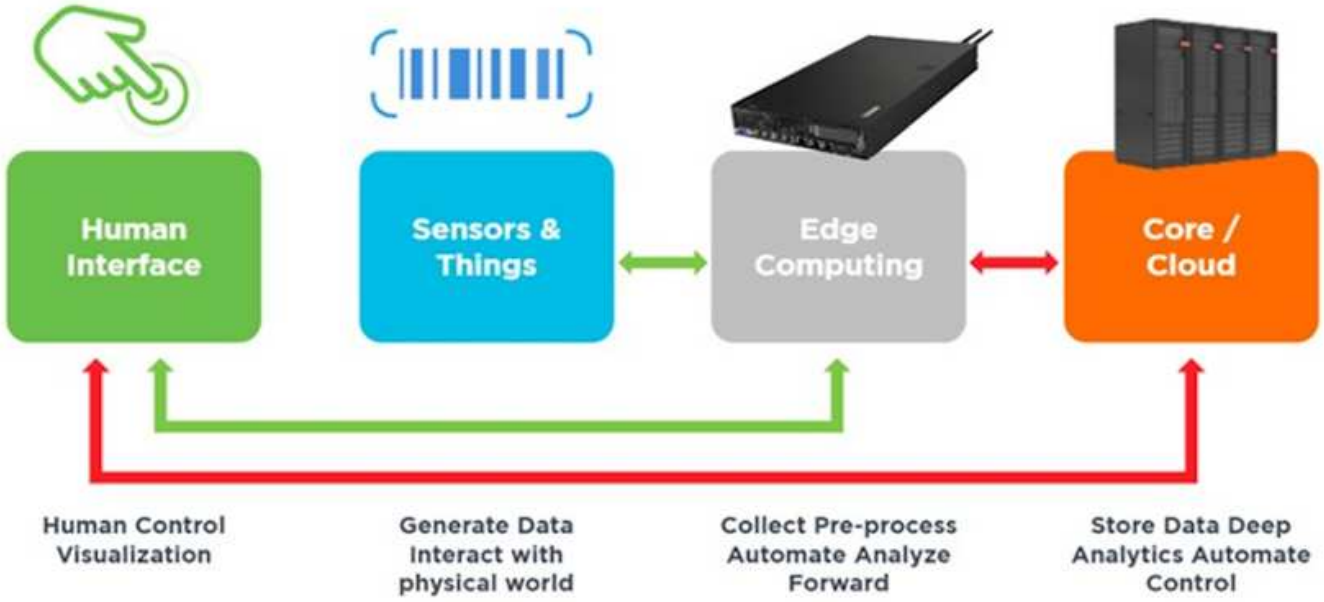
- Diseños modulares y altamente escalables que crecen a medida que lo hace su negocio
- Resiliencia líder en el sector para ahorrar horas de costosos tiempos de inactividad no programados
- Tecnologías flash rápidas para reducir las latencias, acelerar los tiempos de respuesta y gestionar los datos de forma más inteligente en tiempo real

En el ámbito de la IA, Lenovo está adoptando un enfoque práctico para ayudar a las empresas a comprender y adoptar las ventajas DEL APRENDIZAJE AUTOMÁTICO y la IA para sus cargas de trabajo. Los clientes de Lenovo pueden explorar y evaluar las ofertas de IA de Lenovo en los centros de innovación de IA de Lenovo para comprender por completo el valor de su caso de uso en particular. Con el fin de mejorar la rentabilidad de la inversión, este enfoque centrado en el cliente proporciona a los clientes una prueba de concepto para las plataformas de desarrollo de soluciones que están listas para usar y optimizadas para la IA.

Servidor Lenovo ThinkSystem SE350 Edge

La computación perimetral permite analizar los datos de dispositivos de IoT en el extremo de la red antes de enviarlos al centro de datos o al cloud. El sistema ThinkSystem SE350 de Lenovo, como se muestra en la siguiente figura, está diseñado para los requisitos únicos de implementación en el perímetro, con un enfoque en flexibilidad, conectividad, seguridad y capacidad de gestión remota en un factor de forma compacto y reforzado con el medio ambiente.

Con el procesador Intel Xeon D con la flexibilidad necesaria para admitir la aceleración de las cargas de trabajo de IA perimetral, el SE350 se ha diseñado específicamente para afrontar los retos de las implementaciones de servidores en una gran variedad de entornos fuera del centro de datos.



Rendim. MLPerf

MLPerf es el conjunto de pruebas de rendimiento líder del sector para evaluar el rendimiento de la IA. Cubre muchas áreas de IA aplicada, incluida la clasificación de imágenes, la detección de objetos, las imágenes médicas y el procesamiento del lenguaje natural (NLP). En esta validación, hemos utilizado cargas de trabajo de inferencia v0.7, que es la última iteración de la inferencia MLPerf al finalizar esta validación. La "Inferencia del rendimiento ML0,7" la suite incluye cuatro nuevas pruebas de rendimiento para centros de datos y sistemas periféricos:

- **BERT.** representación de encoder bidireccional de Transformers (BERT) ajustada para responder preguntas utilizando el conjunto de datos de escuadrón.
- **DLRM.** el modelo de recomendación de aprendizaje profundo (DLRM) es un modelo de personalización y recomendación que se entrena para optimizar las tarifas de clic (CTR).
- **3D U-Net.** la arquitectura 3D U-Net está entrenada en el conjunto de datos de segmentación del tumor cerebral (Brats).

- **RNN-T.** el transductor de red neural recurrente (RNN-T) es un modelo de reconocimiento automático de voz (ASR) que se entrena en un subconjunto de LibriSpeech. Los resultados y el código de la inferencia de MLPerf están disponibles y se liberan públicamente bajo la licencia de Apache. La inferencia MLPerf tiene una división Edge, que admite los siguientes escenarios:
- **Flujo único.** este escenario imita a los sistemas en los que la capacidad de respuesta es un factor crítico, como las consultas de IA sin conexión realizadas en smartphones. Las consultas individuales se envían al sistema y se registran los tiempos de respuesta. como resultado, se indica la latencia del percentil 90 de todas las respuestas.
- **MultiStream.** este punto de referencia es para sistemas que procesan la entrada de varios sensores. Durante la prueba, las consultas se envían a un intervalo de tiempo fijo. Se impone una limitación de calidad de servicio (latencia máxima permitida). La prueba informa del número de flujos que el sistema puede procesar mientras cumple la restricción QoS.
- **Fuera de línea.** este es el escenario más sencillo que cubre las aplicaciones de procesamiento por lotes y la métrica es el procesamiento en muestras por segundo. Todos los datos están disponibles para el sistema y el punto de referencia mide el tiempo que tarda en procesar todas las muestras.

Lenovo ha publicado puntuaciones de inferencia MLPerf para SE350 con T4, el servidor utilizado en este documento. Consulte los resultados en "https://mlperf.org/inference-results-0-7/" En la sección “Edge, Closed Division”, en la entrada #0.7-145.

Plan de pruebas

Este documento sigue a la inferencia MLPerf v0.7 "codificación", Inferencia MLPerf v1.1 "codificación", y. "bases de datos". Ejecutamos pruebas de rendimiento MLPerf diseñadas para la inferencia en el perímetro tal y como se definen en la tabla siguiente.

Zona	Tarea	Modelo	Conjunto de datos	Tamaño de QSL	Calidad	Limitación de latencia de múltiples flujos
Visión	Clasificación de imágenes	Resnet50v1.5	ImageNET (224 x 224)	1024	99% del FP32	50 ms
Visión	Detección de objetos (grande)	SSD-ResNet34	COCO (1200x1200)	64	99% del FP32	66 ms
Visión	Detección de objetos (pequeño)	SSD-MobileNetsv1	COCO (300 x 300)	256	99% del FP32	50 ms
Visión	Segmentación de imagen médica	UNET 3D	Brats 2019 (224x224x160)	16	99% y 99.9% del FP32	n.a.
Voz	Voz a texto	RNNT	Llibrispeech dev-Clean	2513	99% del FP32	n.a.
Idioma	Procesamiento de idiomas	BERT	Escuadra v1.1	10833	99% del FP32	n.a.

En la siguiente tabla se presentan los escenarios de referencia de Edge.

Zona	Tarea	Escenarios
Visión	Clasificación de imágenes	Flujo único, sin conexión, multisecuencia
Visión	Detección de objetos (grande)	Flujo único, sin conexión, multisecuencia
Visión	Detección de objetos (pequeño)	Flujo único, sin conexión, multisecuencia
Visión	Segmentación de imagen médica	Flujo único, sin conexión
Voz	Voz a texto	Flujo único, sin conexión
Idioma	Procesamiento de idiomas	Flujo único, sin conexión

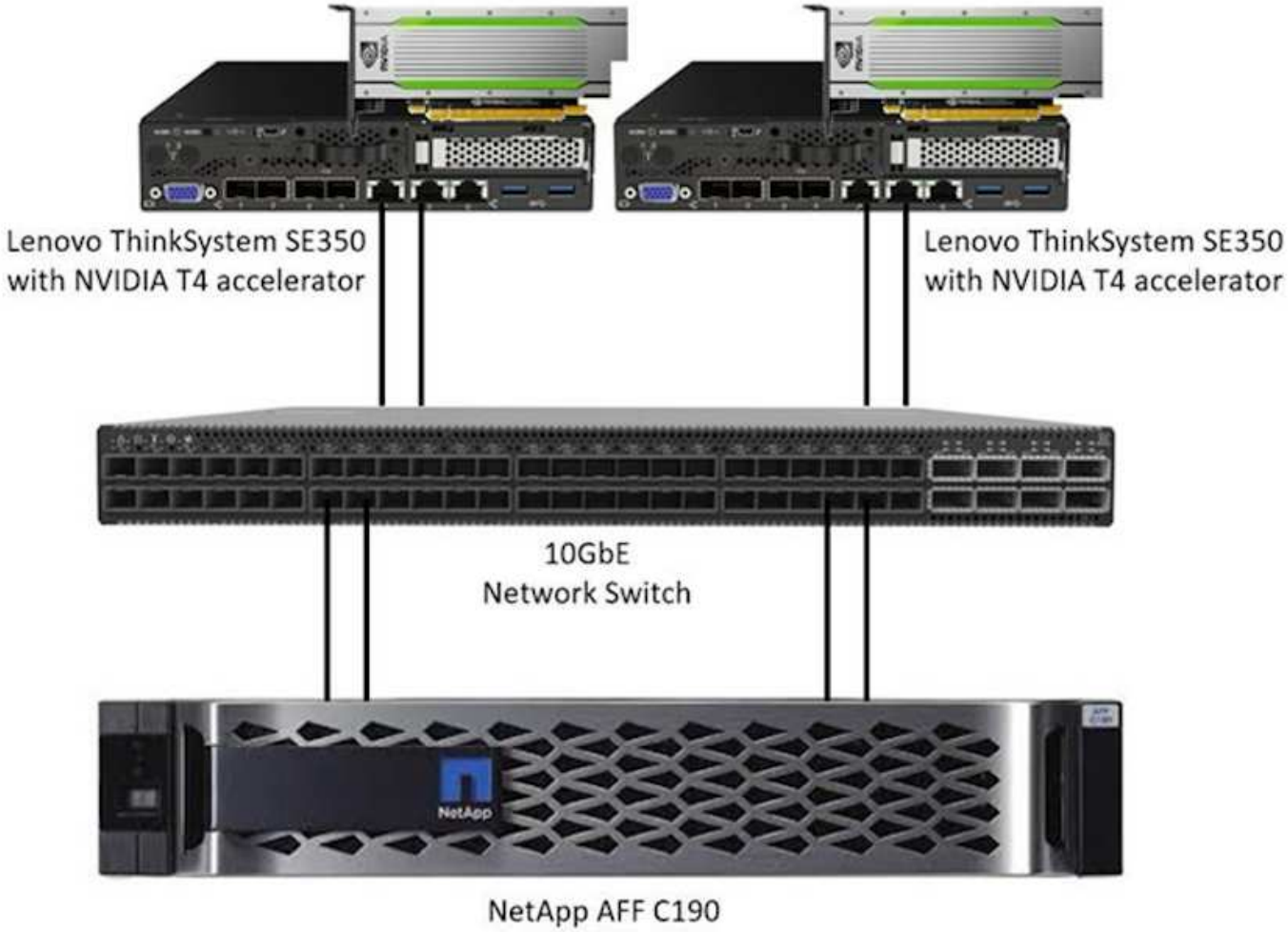
Realizamos estas pruebas de rendimiento utilizando la arquitectura de almacenamiento en red desarrollada en esta validación y comparamos los resultados de las ejecuciones locales en servidores perimetrales que antes se enviaban al MLPerf. La comparación consiste en determinar cuánto impacto tiene el almacenamiento compartido en el rendimiento de la inferencia.

Configuración de prueba

La figura siguiente muestra la configuración de prueba. Utilizamos el sistema de almacenamiento AFF C190 de NetApp y dos servidores Lenovo ThinkSystem SE350 (cada uno con un acelerador NVIDIA T4). Estos componentes se conectan a través de un switch de red de 10 GbE. El almacenamiento en red contiene conjuntos de datos de validación y pruebas y modelos preentrenados. Los servidores proporcionan una funcionalidad computacional y se accede al almacenamiento mediante un protocolo NFS.

En este apartado se describen las configuraciones probadas, la infraestructura de red, el servidor SE350 y los detalles de aprovisionamiento de almacenamiento. En la siguiente tabla se enumeran los componentes básicos para la arquitectura de la solución.

Componentes de la solución	Detalles
Servidores Lenovo ThinkSystem	• 2 servidores SE350 con una tarjeta GPU NVIDIA T4
	• Cada servidor contiene una CPU Intel Xeon D-2123IT con cuatro núcleos físicos que funcionan a 2,20 GHz y 128 GB de RAM
Sistema de almacenamiento AFF de gama básica de NetApp (par de alta disponibilidad)	• Software ONTAP 9 de NetApp • 24 unidades SSD de 960 GB • Protocolo NFS • Un grupo de interfaces por controladora, con cuatro direcciones IP lógicas para puntos de montaje



La siguiente tabla enumera la configuración del almacenamiento: AFF C190 con 2 ranuras de 24 unidades 2RU.

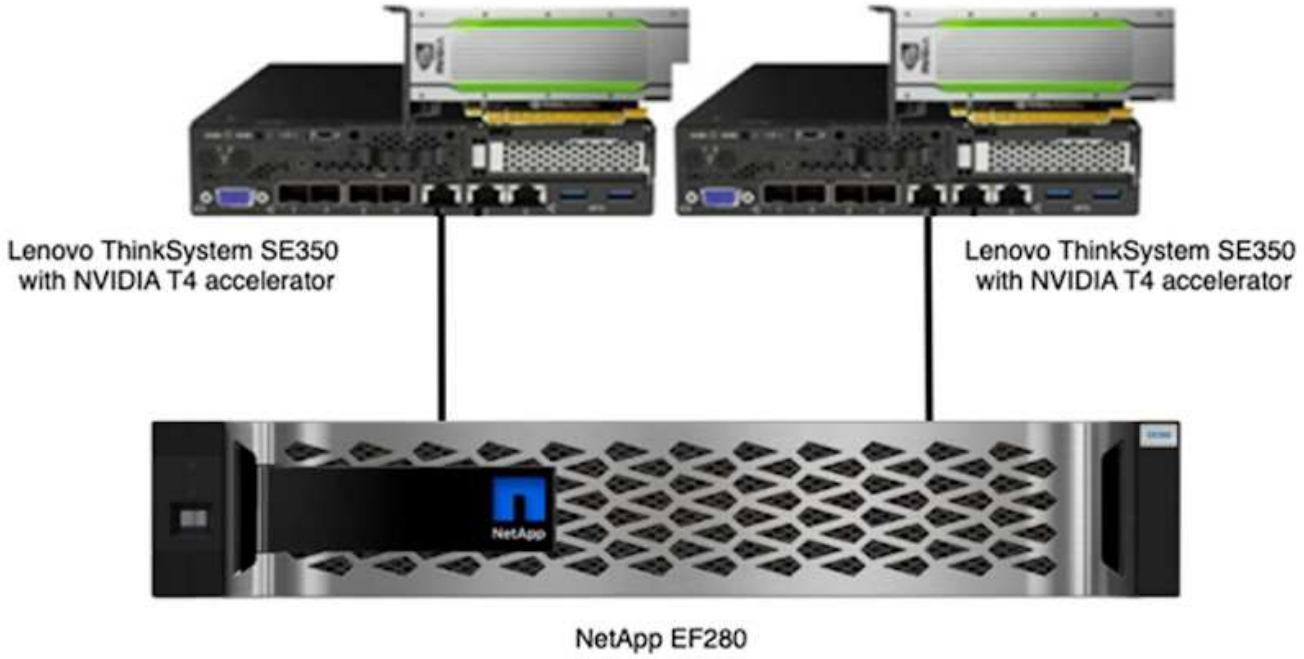
Controladora	Agregado	Volumen FlexGroup	Agregatesize	Volumen	Punto de montaje del sistema operativo
Control1	Agr1	/Netaptenovo_AI_fg	8,42 TIB	15 TB	/netapp_lenovo_fg
Control2	Agr2		8,42 TIB		

La carpeta /netappLenovo_AI_fg contiene los conjuntos de datos utilizados para la validación del modelo.

La siguiente figura muestra la configuración de prueba. Utilizamos el sistema de almacenamiento EF280 de NetApp y dos servidores Lenovo ThinkSystem SE350 (cada uno con un acelerador NVIDIA T4). Estos componentes se conectan a través de un switch de red de 10 GbE. El almacenamiento en red contiene conjuntos de datos de validación y pruebas y modelos preentrenados. Los servidores proporcionan una funcionalidad computacional y se accede al almacenamiento mediante un protocolo NFS.

La siguiente tabla enumera la configuración de almacenamiento para EF280.

Controladora	Grupo de volúmenes	Volumen	Volumen	Tamaño DDPsize	Método de conexión
Control1	DDP1	Volumen 1	8,42 TIB	16 TB	SE350-1 a iSCSI LUN 0
Control2		Volumen 2	8,42 TIB		SE350-2 a iSCSI LUN 1



Procedimiento de prueba

Este apartado describe los procedimientos de prueba utilizados para validar esta solución.

Configuración de inferencia de IA y sistema operativo

Para AFF C190, utilizamos Ubuntu 18.04 con controladores NVIDIA y docker con soporte para GPU de NVIDIA y usamos MLPerf "codificación" Disponible como parte de la presentación de Lenovo a la inferencia MLPerf v0.7.

Para EF280, utilizamos Ubuntu 20.04 con controladores NVIDIA y docker con soporte para las GPU de NVIDIA y MLPerf "codificación" Disponible como parte de la presentación de Lenovo a la inferencia MLPerf v1.1.

Para configurar la inferencia de IA, siga estos pasos:

1. Descargue los conjuntos de datos que requieren registro, el conjunto de datos ImageNET 2012 Validation, el conjunto de datos Criteo Terabyte y el conjunto de entrenamiento Brats 2019 y, a continuación, descomprima los archivos.
2. Cree un directorio de trabajo con al menos 1 TB y defina una variable ambiental `MLPERF_SCRATCH_PATH`

referencia al directorio.

Debe compartir este directorio en el almacenamiento compartido para el caso de uso del almacenamiento de red o en el disco local cuando realice pruebas con datos locales.

3. Ejecute la Marca `prebuild` comando, que crea e inicia el contenedor docker para las tareas de inferencia necesarias.

   Los siguientes comandos se ejecutan desde el contenedor docker en ejecución:

   - Descargue modelos de IA preformados para tareas de inferencia de MLPerf: `make download_model`
   - Descargue conjuntos de datos adicionales que se pueden descargar gratuitamente: `make download_data`
   - Preprocesar los datos: Make `preprocess_data`
   - Ejecución: `make build`.
   - Cree motores de inferencia optimizados para la GPU en servidores informáticos: `make generate_engines`
   - Para ejecutar cargas de trabajo de inferencia, ejecute el siguiente (un comando):

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

Se ejecuta la inferencia de IA

Se ejecutaron tres tipos de ejecuciones:

- Inferencia de IA de un único servidor con almacenamiento local
- Inferencia de IA de un único servidor con el almacenamiento en red
- Inferencia de la IA en varios servidores con el almacenamiento en red

Resultados de la prueba

Se llevaron a cabo una multitud de pruebas para evaluar el rendimiento de la arquitectura propuesta.

Existen seis cargas de trabajo diferentes (clasificación de imágenes, detección de objetos [pequeño], detección de objetos [grande], imágenes médicas, voz a texto, Y procesamiento de lenguaje natural [NLP]), que se puede ejecutar en tres escenarios diferentes: Sin conexión, flujo único y multisecuencia.

El último escenario se implementa sólo para la clasificación de imágenes y la detección de objetos.

Esto permite 15 cargas de trabajo posibles, todas ellas probadas en tres configuraciones diferentes:

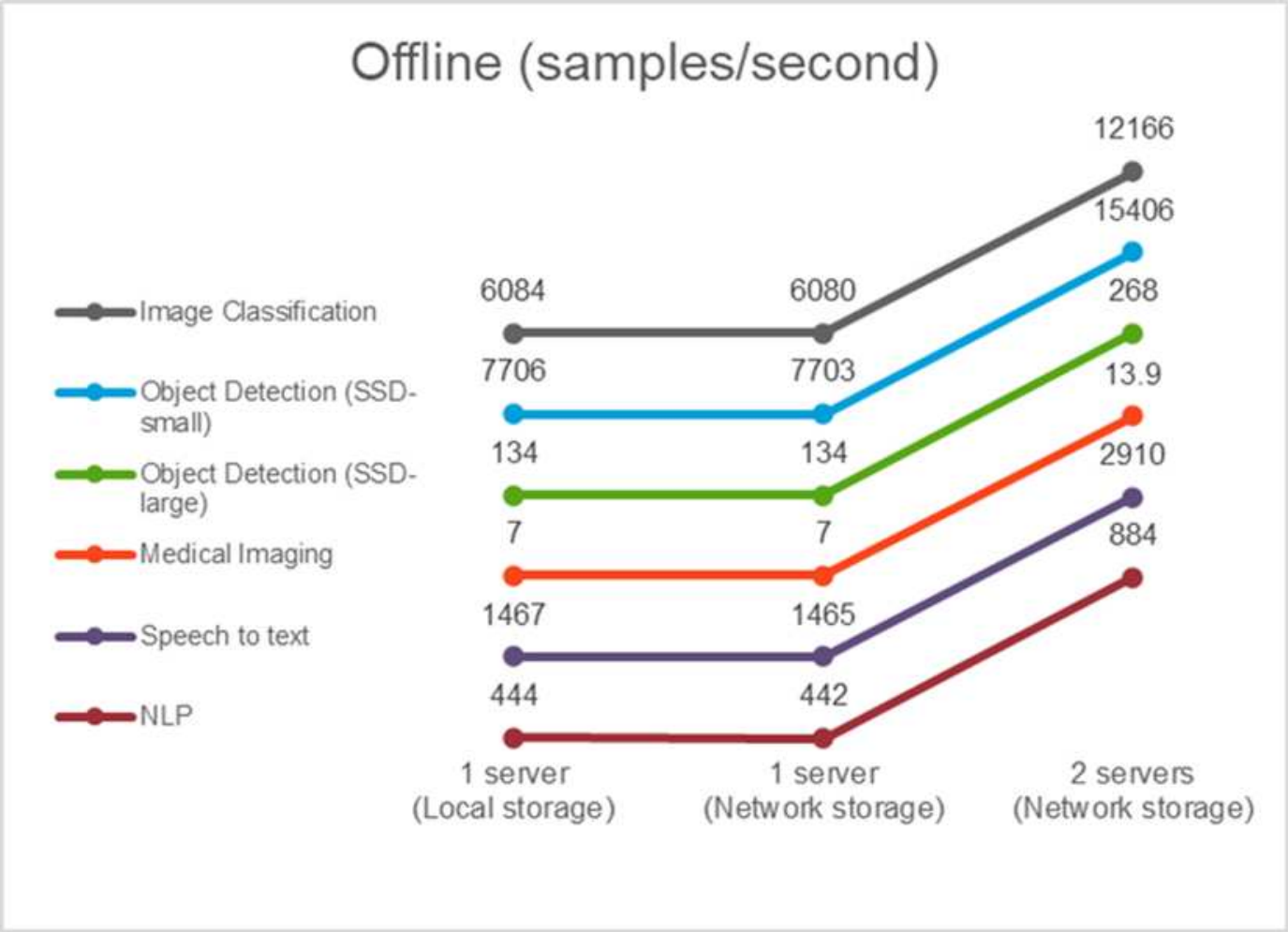
- Un único servidor/almacenamiento local
- Almacenamiento único de servidor/red

- Almacenamiento en red/multiservidor

Los resultados se describen en las siguientes secciones.

Inferencia de la IA en un escenario sin conexión para AFF

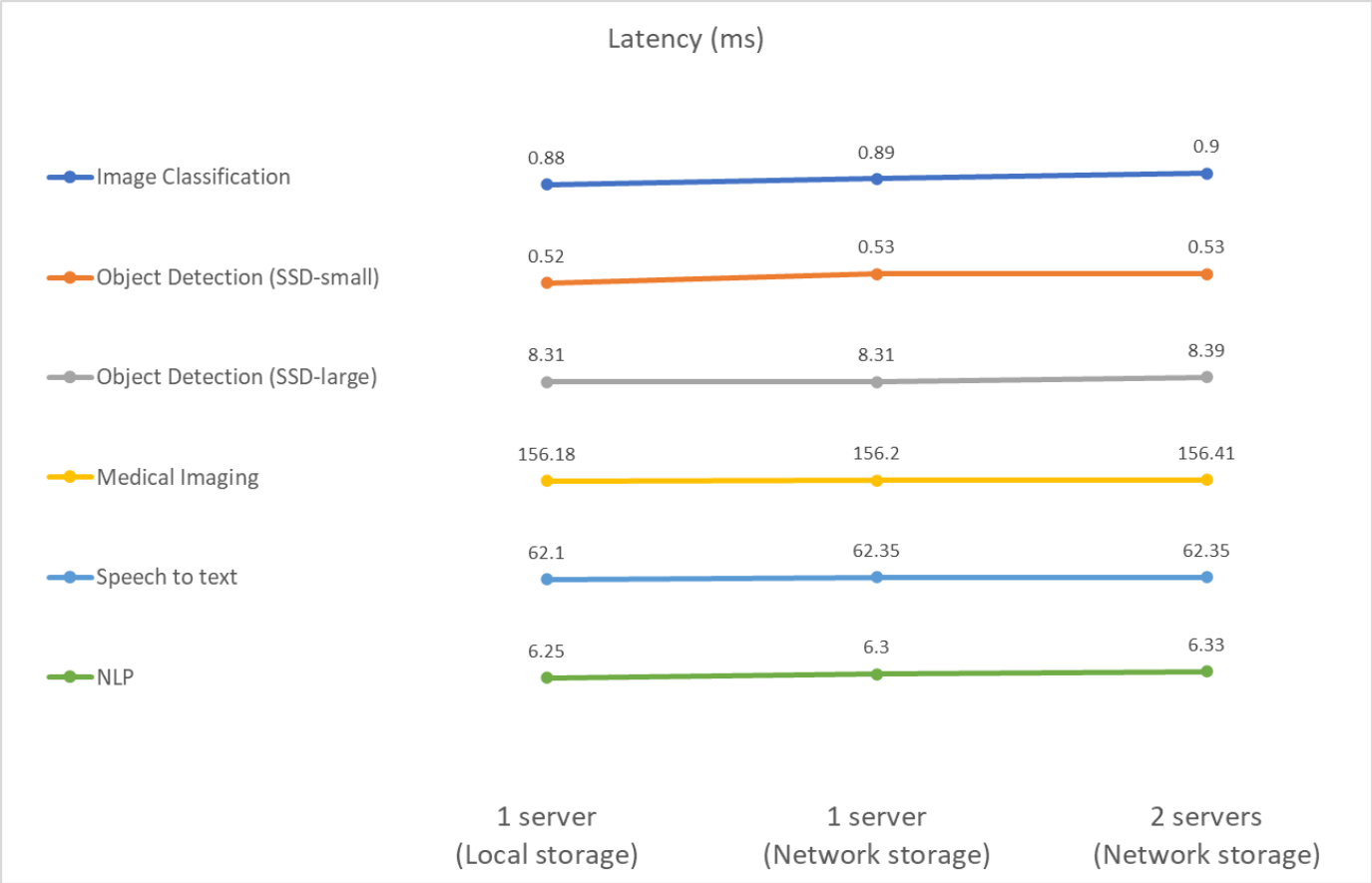
En este escenario, todos los datos estaban disponibles para el servidor y se midió el tiempo que se tardaba en procesar todas las muestras. Reportamos los anchos de banda en muestras por segundo como resultados de las pruebas. Cuando se utilizaron más de un servidor informático, hemos indicado el total de ancho de banda total en todos los servidores. En la siguiente figura se muestran los resultados de los tres casos de uso. Para el caso de dos servidores, reportamos un ancho de banda combinado de ambos servidores.



Los resultados muestran que el almacenamiento en red no afecta de manera negativa al rendimiento; el cambio es mínimo y, en algunas tareas, no se encuentra ninguno. Cuando se añade el segundo servidor, el ancho de banda total se duplica exactamente o, en el peor de los casos, el cambio es inferior al 1%.

Inferencia de IA en un escenario de flujo único para AFF

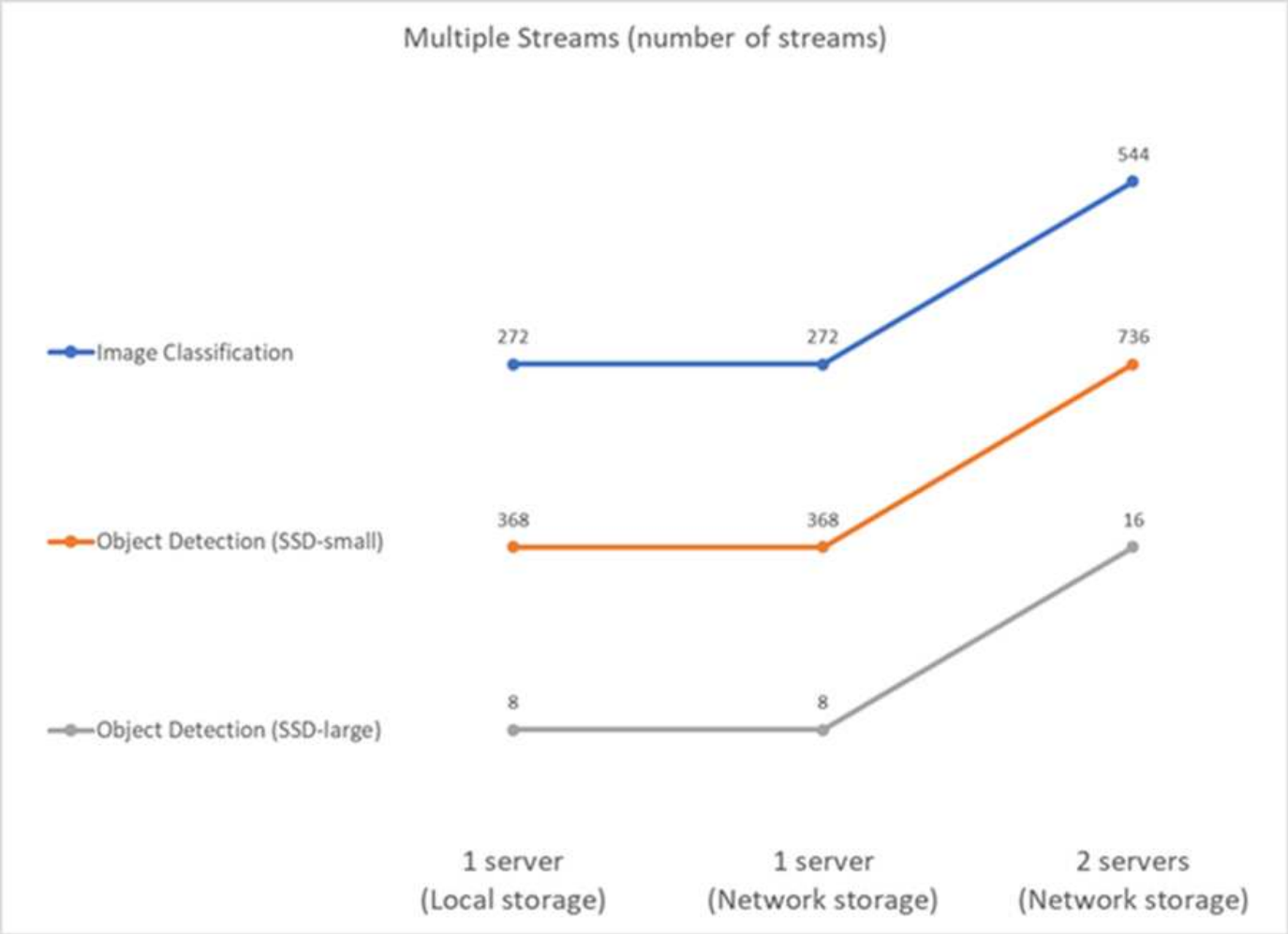
Esta prueba mide la latencia. Para el caso de varios servidores informáticos, reportamos la latencia media. Los resultados del conjunto de tareas se indican en la siguiente figura. Para el caso de dos servidores, hemos registrado la latencia media de ambos servidores.



Los resultados, de nuevo, muestran que el almacenamiento de red es suficiente para manejar las tareas. La diferencia entre el almacenamiento local y el de red en un caso de servidor es mínima o ninguna. De igual modo, cuando dos servidores utilizan el mismo almacenamiento, la latencia de ambos servidores se mantiene igual o se modifica en una cantidad muy pequeña.

Inferencia de IA en un escenario de transmisión múltiple para AFF

En este caso, el resultado es el número de flujos que el sistema puede manejar mientras se satisface la restricción QoS. Por lo tanto, el resultado siempre es un entero. En más de un servidor, reportamos el número total de flujos sumados en todos los servidores. No todas las cargas de trabajo admiten este escenario, pero hemos ejecutado las que lo hacen. Los resultados de nuestras pruebas se resumen en la siguiente figura. Para el caso de dos servidores, reportamos el número combinado de flujos de ambos servidores.



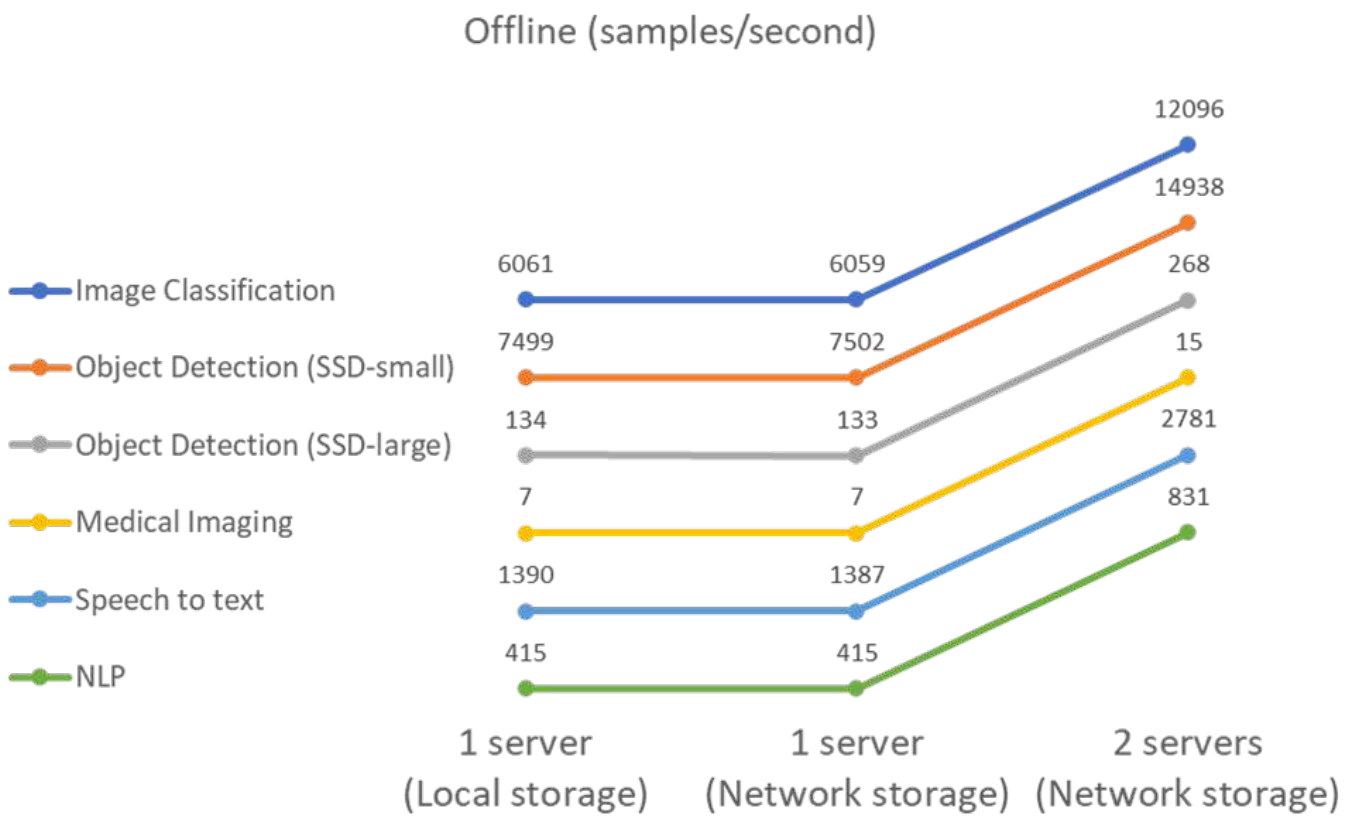
Los resultados muestran un rendimiento perfecto de la configuración; el almacenamiento local y en red ofrecen los mismos resultados y al añadir el segundo servidor se duplica el número de flujos que puede gestionar la configuración propuesta.

Resultados de la prueba para EF

Se llevaron a cabo una multitud de pruebas para evaluar el rendimiento de la arquitectura propuesta. Existen seis cargas de trabajo diferentes (clasificación de imágenes, detección de objetos [pequeño], detección de objetos [grande], imágenes médicas, voz a texto, Y procesamiento de lenguaje natural [NLP]), que se ejecutaron en dos escenarios diferentes: Offline y de flujo único. Los resultados se describen en las siguientes secciones.

Inferencia de la IA en un escenario sin conexión para EF

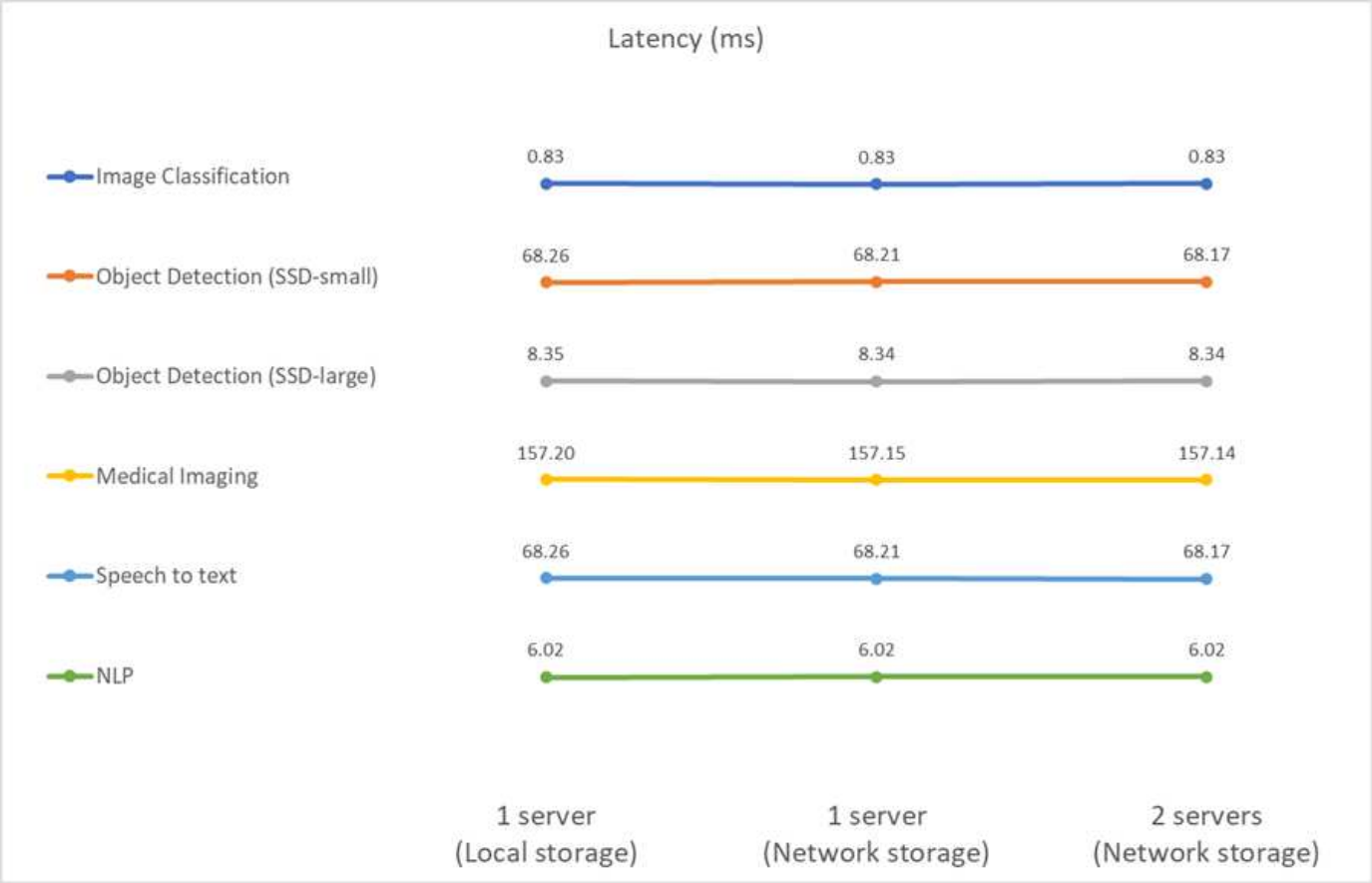
En este escenario, todos los datos estaban disponibles para el servidor y se midió el tiempo que se tardaba en procesar todas las muestras. Reportamos los anchos de banda en muestras por segundo como resultados de las pruebas. En las ejecuciones de un solo nodo reportamos la media de ambos servidores, mientras que en dos ejecuciones de servidor reportamos un ancho de banda total resumido en todos los servidores. Los resultados de los casos de uso se muestran en la siguiente figura.



Los resultados muestran que el almacenamiento en red no afecta de manera negativa al rendimiento; el cambio es mínimo y, en algunas tareas, no se encuentra ninguno. Cuando se añade el segundo servidor, el ancho de banda total se duplica exactamente o, en el peor de los casos, el cambio es inferior al 1%.

Inferencia de IA en un escenario de flujo único para EF

Esta prueba mide la latencia. En todos los casos, reportamos una latencia media en todos los servidores involucrados en las ejecuciones. Los resultados para el conjunto de tareas se dan.



Los resultados muestran de nuevo que el almacenamiento en red es suficiente para manejar las tareas. La diferencia entre el almacenamiento local y el de red en un caso de servidor es mínima o ninguna. De igual modo, cuando dos servidores utilizan el mismo almacenamiento, la latencia de ambos servidores se mantiene igual o se modifica en una cantidad muy pequeña.

Opciones de ajuste de tamaño de arquitectura

Puede ajustar la configuración utilizada para la validación y adaptarla a otros casos prácticos.

Servidor de computación

Utilizamos una CPU Intel Xeon D-2123IT, que es el nivel más bajo de CPU compatible con SE350, con cuatro núcleos físicos y 60 W TDP. Aunque el servidor no admite la sustitución de CPU, se puede pedir con una CPU más potente. La CPU más alta admitida es Intel Xeon D-2183IT con 16 núcleos, 100 W con 2,20 GHz. Esto aumenta considerablemente la capacidad computacional de la CPU. Mientras que la CPU no era un cuello de botella para ejecutar las cargas de trabajo de inferencia en sí, ayuda con el procesamiento de datos y otras tareas relacionadas con la inferencia. Actualmente, NVIDIA T4 es la única GPU disponible para casos de uso periféricos; por lo tanto, actualmente, no es posible actualizar o degradar el GPU.

Almacenamiento compartido

Para las pruebas y la validación, se utilizó el sistema AFF C190 de NetApp, que tiene una capacidad de almacenamiento máxima de 50,5 TB, una rendimiento de 4,4 GB/s para lecturas secuenciales y 230 000 IOPS para pequeñas lecturas aleatorias, con los fines de este documento y se ha demostrado ser apto para cargas de trabajo de inferencia periférica.

No obstante, si se requiere más capacidad de almacenamiento o velocidades de red más rápidas, deberá utilizar el o el sistema AFF A220 de NetApp "AFF A250 de NetApp" sistemas de almacenamiento. Además, el sistema EF280 de NetApp, que tiene una capacidad máxima de 1,5 PB, 10 Gbps de ancho de banda, también se utilizó para la validación de esta solución. Si prefiere más capacidad de almacenamiento con un mayor ancho de banda, "EF300 de NetApp" puede utilizarse.

Conclusión

La automatización impulsada por IA y la informática perimetral es un enfoque líder que ayuda a las organizaciones empresariales a lograr la transformación digital y maximizar la eficiencia y la seguridad operativas. En el caso de los entornos periféricos, los datos se procesan con mayor rapidez, porque no tienen que desplazarse hacia y desde un centro de datos. Por lo tanto, el coste asociado al envío de datos a los centros de datos o al cloud se ve disminuido. La reducción de la latencia y el aumento de la velocidad pueden ser beneficiosas cuando las empresas deben tomar decisiones casi en tiempo real mediante modelos de inferencia de IA implementados en el perímetro.

Los sistemas de almacenamiento de NetApp ofrecen el mismo rendimiento o mejor que el almacenamiento SSD local y ofrecen las siguientes ventajas a los científicos de datos, ingenieros de datos, desarrolladores de IA/ML y responsables de la toma de decisiones EMPRESARIALES o TECNOLÓGICAS:

- Uso compartido de datos sin esfuerzo entre sistemas de IA, análisis y otros sistemas de negocio cruciales. Este uso compartido de datos reduce la sobrecarga de la infraestructura, mejora el rendimiento y optimiza la gestión de datos en toda la empresa.
- Almacenamiento y cálculo escalables de forma independiente para minimizar los costes y mejorar el uso de recursos.
- Flujos de trabajo de desarrollo y puesta en marcha optimizados que utilizan copias Snapshot integradas y clones para espacios de trabajo de usuario instantáneos con gestión eficiente del espacio, control de versiones integrado y una puesta en marcha automatizada.
- Protección de datos de clase empresarial para la recuperación ante desastres y la continuidad del negocio. La solución de NetApp y Lenovo presentada en este documento es una arquitectura flexible y de escalado horizontal ideal para puestas en marcha de inferencia de IA en el perímetro.

Reconocimientos

- J.J. Falkanger, Sr. Director de Soluciones HPC e IA de Lenovo
- Dave Arnette, ingeniero técnico de marketing, NetApp
- Joey Parnell, responsable técnico de soluciones de IA de E-Series de NetApp
- Cody Harryman, ingeniero de control de calidad, NetApp

Dónde encontrar información adicional

Para obtener más información sobre la información descrita en este documento, consulte los siguientes documentos y/o sitios web:

- Página de producto de las cabinas AFF a-Series de NetApp

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

- Software de gestión de datos ONTAP de NetApp: Biblioteca de información de ONTAP 9

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

- TR-4727: Introducción a EF-Series de NetApp

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

- Especificaciones técnicas del software SANtricity para E-Series de NetApp

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

- Almacenamiento persistente de NetApp para contenedores: Trident de NetApp

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

- Rendim. MLPerf
  - "https://mlcommons.org/en/"
  - "http://www.image-net.org/"
  - "https://mlcommons.org/en/news/mlperf-inference-v11/"
- Copia y sincronización de NetApp BlueXP

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

- Prueba de rendimiento de TensorFlow

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

- Servidor Lenovo ThinkSystem SE350 Edge

  "https://lenovopress.com/lp1168"

- Sistema de almacenamiento flash unificado Lenovo ThinkSystem DM5100F

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

Información de copyright

Copyright © 2024 NetApp, Inc. Todos los derechos reservados. Imprimido en EE. UU. No se puede reproducir este documento protegido por copyright ni parte del mismo de ninguna forma ni por ningún medio (gráfico, electrónico o mecánico, incluidas fotocopias, grabaciones o almacenamiento en un sistema de recuperación electrónico) sin la autorización previa y por escrito del propietario del copyright.

El software derivado del material de NetApp con copyright está sujeto a la siguiente licencia y exención de responsabilidad:

ESTE SOFTWARE LO PROPORCIONA NETAPP «TAL CUAL» Y SIN NINGUNA GARANTÍA EXPRESA O IMPLÍCITA, INCLUYENDO, SIN LIMITAR, LAS GARANTÍAS IMPLÍCITAS DE COMERCIALIZACIÓN O IDONEIDAD PARA UN FIN CONCRETO, CUYA RESPONSABILIDAD QUEDA EXIMIDA POR EL PRESENTE DOCUMENTO. EN NINGÚN CASO NETAPP SERÁ RESPONSABLE DE NINGÚN DAÑO DIRECTO, INDIRECTO, ESPECIAL, EJEMPLAR O RESULTANTE (INCLUYENDO, ENTRE OTROS, LA OBTENCIÓN DE BIENES O SERVICIOS SUSTITUTIVOS, PÉRDIDA DE USO, DE DATOS O DE BENEFICIOS, O INTERRUPCIÓN DE LA ACTIVIDAD EMPRESARIAL) CUALQUIERA SEA EL MODO EN EL QUE SE PRODUJERON Y LA TEORÍA DE RESPONSABILIDAD QUE SE APLIQUE, YA SEA EN CONTRATO, RESPONSABILIDAD OBJETIVA O AGRAVIO (INCLUIDA LA NEGLIGENCIA U OTRO TIPO), QUE SURJAN DE ALGÚN MODO DEL USO DE ESTE SOFTWARE, INCLUSO SI HUBIEREN SIDO ADVERTIDOS DE LA POSIBILIDAD DE TALES DAÑOS.

NetApp se reserva el derecho de modificar cualquiera de los productos aquí descritos en cualquier momento y sin aviso previo. NetApp no asume ningún tipo de responsabilidad que surja del uso de los productos aquí descritos, excepto aquello expresamente acordado por escrito por parte de NetApp. El uso o adquisición de este producto no lleva implícita ninguna licencia con derechos de patente, de marcas comerciales o cualquier otro derecho de propiedad intelectual de NetApp.

Es posible que el producto que se describe en este manual esté protegido por una o más patentes de EE. UU., patentes extranjeras o solicitudes pendientes.

LEYENDA DE DERECHOS LIMITADOS: el uso, la copia o la divulgación por parte del gobierno están sujetos a las restricciones establecidas en el subpárrafo (b)(3) de los derechos de datos técnicos y productos no comerciales de DFARS 252.227-7013 (FEB de 2014) y FAR 52.227-19 (DIC de 2007).

Los datos aquí contenidos pertenecen a un producto comercial o servicio comercial (como se define en FAR 2.101) y son propiedad de NetApp, Inc. Todos los datos técnicos y el software informático de NetApp que se proporcionan en este Acuerdo tienen una naturaleza comercial y se han desarrollado exclusivamente con fondos privados. El Gobierno de EE. UU. tiene una licencia limitada, irrevocable, no exclusiva, no transferible, no sublicenciable y de alcance mundial para utilizar los Datos en relación con el contrato del Gobierno de los Estados Unidos bajo el cual se proporcionaron los Datos. Excepto que aquí se disponga lo contrario, los Datos no se pueden utilizar, desvelar, reproducir, modificar, interpretar o mostrar sin la previa aprobación por escrito de NetApp, Inc. Los derechos de licencia del Gobierno de los Estados Unidos de América y su Departamento de Defensa se limitan a los derechos identificados en la cláusula 252.227-7015(b) de la sección DFARS (FEB de 2014).

Información de la marca comercial

NETAPP, el logotipo de NETAPP y las marcas que constan en http://www.netapp.com/TM son marcas comerciales de NetApp, Inc. El resto de nombres de empresa y de producto pueden ser marcas comerciales de sus respectivos propietarios.