# NetApp AI and Lenovo ThinkSystem for AI Inferencing

Operationalize your AI
workloads at the edge

**█ NetApp** | **Lenovo**

Modern applications that are driven by artificial intelligence (AI) and machine learning (ML) are straining the limits of cloud computing. End users and devices demand access to applications, data, and services at any place and any time, with minimal latency. To meet these demands, resources are being moved to the data sources, enabling the lower latency required by real time analytics. This migration reduces the costs associated with data storage and transfer.

In the context of AI, the core data center is a platform that provides centralized services, such as machine learning and analytics, and the edge is where the real-time production data is subject to inferencing. These edge deployments are usually connected to a core data center. They provide end-user services and serve as a staging layer for data generated by IoT devices that needs additional processing and that is too time sensitive to be transmitted back to a centralized core.

### The solution
NetApp and Lenovo have partnered to develop an affordable, easy to manage edge inferencing solution. This validated edge solution delivers a simple, smart, and secure solution at an affordable price. It helps you meet requirements with a modern all-flash array that provides comprehensive data services, integrated data protection, seamless scalability, new levels of performance, and cloud integration.

### Affordable performance with maximum data protection
#### Speed inference and get to insights faster
The Lenovo® ThinkSystem™ SE350 an edge server designed to enable traditional IT and OT applications as well as new transformative IoT and AI systems. The ThinkSystem SE350 is built on the Intel Xeon D2100 CPU, and is a compact, rugged system designed to go into any environment. The NetApp® AFF C190 system is optimized for flash and delivers 10 times faster application response than hybrid arrays. All-flash storage allows you to run more workloads on a single system without compromising performance.

**Key benefits**

**Affordable performance with maximum data protection**
- Optimized for flash at the lowest possible cost.
- Lenovo SE350 leads the latest round of MLPerf inference benchmarks for edge servers.
- Keep your data safe with built-in data protection capabilities such as volume-level encryption and NetApp Snapshot copies.

**Streamline operations and reduce TCO**
- Reduce the amount of storage needed with NetApp data compaction and deduplication, and lower storage costs with automatic cold data tiering.
- Independently scale compute and storage to minimize costs and improve resource utilization.
- Simplify management with a unified management interface and a single set of tools.

**Future-proof your infrastructure**
- Meet demanding and constantly changing business needs with a solution that delivers seamless scalability, easy cloud connectivity, and integration with emerging applications.
- Deploy a proven enterprise-grade, nondisruptive scalable solution for organizations of all sizes.

The design in Figure 1 shows how multiple ThinkSystem SE350s can be deployed in an edge environmen (e.g., multiple retail stores). The model management can be handled through a single storage node (AFF C190) and pushed out to each of the compute nodes, allowing you to simplify model management for your AI workloads. This design also provides local data storage so that you do not have to move all of the data from the compact edge servers back to the cloud. This tiering allows you to reduce your storage cost by retaining the data locally and only moving necessary data back to the cloud.
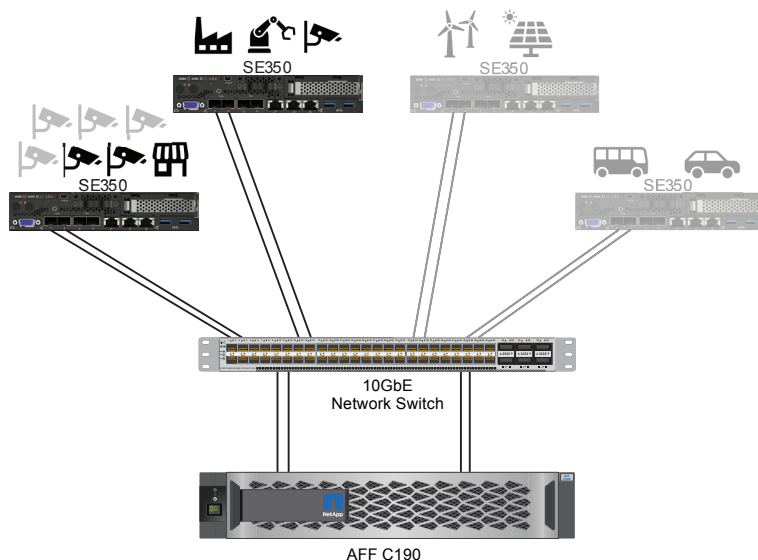
Figure 1: NetApp and Lenovo validated inferencing solution architecture.

- **Lenovo SE350**
  - Edge server–small footprint, low power
  - 1 NVIDIA T4 computational accelerator

- **NetApp® C190**
  - Entry-level AFF system
  - High-end data management

- **Test workload**
  - MLPerf v0.7 inference benchmark suite

## Protect your data

Your end users demand nonstop access to business-critical data, so you can't allow unplanned downtime and other disruptions to affect business operations. The AFF C190 system offers  comprehensive integrated data protection to safeguard your data and keep things running smoothly. Meet your requirements for local backup with near-instant recovery by using space-efficient NetApp Snapshot™ copies. Achieve remote backup/recovery and disaster recovery with NetApp SnapMirror® asynchronous replication.

## Streamline operations and reduce TCO
### Simplify data management

Data management is crucial to ensure that appropriate resources are used for applications and for datasets. NetApp ONTAP® 9, the industry-leading enterprise data management software, simplifies management of the most complex environments. A single management interface with a single set of tools streamlines and simplifies operations regardless of where your data resides. It also allows you to freely move your data to wherever it's needed—the edge, the core, or the cloud.

### Lower Infrastructure costs

Reduce infrastructure costs with the comprehensive storage efficiency capabilities delivered by the AFF C190 system. Data reduction features in ONTAP, including compression and deduplication, work together to help you optimize the capacity needed.

Your storage efficiency is backed by the industry-leading efficiency guarantee program.

## Secure your data

Robust security is important to protect data from internal and external threats. Without any special hardware, you can easily secure your at-rest data with NetApp Volume Encryption. In addition, multifactor authentication improves data access control, and cryptographic shredding enables General Data Protection Regulation (GDPR) compliance for your organization.

The Lenovo SE350 is equipped with ThinkShield security featuring cybersecurity capabilities with key encrypted storage and secured bios, and physical security capabilities such as a locking bezel and intrusion and tamper protection mechanisms.

## Future-proof your infrastructure

NetApp ONTAP is the most cloud-connected storage management software, with options for software-defined storage (ONTAP Select) and cloud-native instances (NetApp Cloud Volumes Service) in all public clouds. To accommodate your ever-expanding data growth, ONTAP supports the nondisruptive addition of capacity to existing controllers as well as to scale-out clusters, and it offers enterprise-grade data services for next-generation platforms and applications. And with multiple Lenovo servers, you get full scalability of data storage and computation power.

**About Lenovo**

Lenovo (HKSE: 992) (ADR: LNVGY) is a US$50 billion Fortune Global 500 company, with 63,000 employees and operating in 180 markets around the world. Focused on a bold vision to deliver smarter technology for all, we are developing world-changing technologies that create a more inclusive, trustworthy and sustainable digital society. By designing, engineering and building the world's most complete portfolio of smart devices and infrastructure, we are also leading an Intelligent Transformation – to create better experiences and opportunities for millions of customers around the world. To find out more visit https://www.lenovo.com, follow us on LinkedIn, Facebook, Twitter, YouTube, Instagram, Weibo and read about the latest news via our StoryHub.

**About NetApp**

In a world full of generalists, NetApp is a specialist. We're focused on one thing, helping your business get the most out of your data. NetApp brings the enterprise-grade data services you rely on into the cloud, and the simple flexibility of cloud into the data center. Our industry-leading solutions work across diverse customer environments and the world's biggest public clouds.

As a cloud-led, data-centric software company, only NetApp can help build your unique data fabric, simplify and connect your cloud, and securely deliver the right data, services and applications to the right people—anytime, anywhere. www.netapp.com

**NetApp**