DELLTechnologies

NVIDIA

Accelerating the Future with AI and Edge Computing

Get started >

Transform Your Organization by Capitalizing on Data at the Edge

Data fuels today's business world

Every transaction, sensor reading, and customer interaction generates and utilizes immense amounts of information. Yet organizations too often are unable to act on this data quickly or effectively enough to unlock its full value. The solution lies at the edge—a space where data is generated, and where actions happen as fast as the data flows. AI and edge computing are expanding the boundaries of what's possible in this data-driven era, revolutionizing industries and reshaping how we live and work.

Edge AI is the next wave of innovation, and those who act now will lead their industries into a brighter, more intelligent future.

Contents

Speed: Accelerate Action from Data Processed at its Source

Bring AI to your data

The world creates an overwhelming influx of data every second. According to Gartner, **75% of enterprise-managed data** will be generated at the edge in **2025**,[1] and 451 Research predicts that over **60% of data-intensive compute** will reside in edge environments by **2027**.[2] Too often, this data sits untapped as many IT teams often struggle with the speed, volume and requirements of real-time data processing throughout edge environments.

AI and edge computing offer the real-time intelligence needed to act upon this data explosion and unlock new realms of innovation. From facilitating data-driven decision-making to enabling smarter operations, industries that adapt to this evolution will lead in a world where agility and insight define success.

This is where edge computing redefines the landscape. By bringing edge-optimized compute to where data is generated, and leveraging features such as zero-touch provisioning and automated application orchestration, you can eliminate latency caused by transmission of data to cloud environments and core data centers. Decisions can be made and actions can be taken in milliseconds, not hours. AI amplifies this further, providing predictive insights, automation, and intelligence exactly where it's needed most.

[1] Gartner, "Innovation Insight for Edge AI," Arun Chandrasekaran & Eric Goodness, April 10, 2024

[2] 451 Research Market Monitor analysis of low-latency workloads across eight key industries in the US; Edge Workload Total Addressable Market (TAM) Analysis, 2023.

Scale: Power the Evolving Requirements of AI Inferencing

The shift to edge AI is necessary for organizations looking to overcome increasing complexity as operations expand throughout diverse geographies and industries. This places burden on traditional, centralized infrastructures, as scaling to meet unique requirements at each location becomes costly and time-consuming. Managing distributed environments while maintaining efficiency and consistency requires a new approach.

Edge AI requires scalable solutions that enable organizations to deploy intelligent systems closer to where decisions are made. This flexibility ensures operations in factories, remote sites, and retail locations remain synchronized and effective, regardless of workload placement and geographical diversity. Centralized management of edge infrastructure and applications optimizes workflows, while adapting seamlessly to handle growth and changing workload requirements in real-time.

Edge AI empowers organizations to replicate successful models across multiple sites without re-engineering core systems. Whether customizing production lines in manufacturing or supporting personalized customer experiences globally, edge AI delivers scalable intelligence tailored to varying needs – backed by a comprehensive ecosystem of software solutions. By planning for scalability upfront, businesses are positioning themselves to adapt and thrive in dynamic markets with unparalleled agility.

Security: Protect Data and Enhance Reliability Across Dispersed Environments

Handling sensitive data that can inspire innovative processes demands uncompromising security. While distributed edge environments can introduce vulnerabilities, edge inferencing safeguards any sensitive data by processing and acting on data at its source. This reduces reliance on centralized systems, minimizing exposure to cyber threats and unauthorized access while enhancing privacy. Sensitive information often remains contained in an air-gapped vault, eliminating the risk associated with transmission to a core data center or cloud environment.

Edge inferencing also simplifies regulatory compliance by ensuring data stays within localized environments. Whether adhering to privacy laws like GDPR or industry-specific audit requirements, organizations can meet stringent standards and align with data sovereignty guidelines without compromising efficiency. By integrating robust encryption and zero trust protocols, organizations can build a resilient edge that unlocks the potential of real-time intelligence while protecting critical assets.

According to IDC, 75% of organizations say the workloads they run at the edge are highly critical or critical.[3]

[3] IDC EdgeView 2024, n=800

How to Take Advantage of Edge AI Across Industries

Energy & Utilities

Deploy intelligent energy distribution and grid management powered by Edge AI. From predicting peak demand for better grid stability to optimizing renewable energy integration, edge AI processes real-time data to create more efficient and resilisient energy systems.

Financial Services

Enhance fraud detection and risk management measures with edge AI. By analyzing transactional data locally, financial institutions can respond to threats in milliseconds while maintaining regulatory compliance.

Healthcare

Respond to critical medical scenarios by deploying diagnostic tools and real-time patient monitoring at the point of care. Edge AI ensures immediate insights and faster decision-making during emergencies, improving outcomes and saving lives.

Manufacturing

Deliver real-time quality control using edge AI solutions that analyze production data instantly. Proactively identify equipment failures with precision using predictive maintenance, minimizing downtime and optimizing workflows—all without relying on centralized systems.

Retail

Personalize shopping experiences by analyzing customer behavior in-store in real-time. Edge AI streamlines operations, enabling dynamic inventory management and enhancing efficiency at physical locations.

Smart Cities

Implement adaptive infrastructure and traffic management powered by edge AI. From coordinating traffic lights for smoother flow to optimizing energy use in urban systems, edge AI analyzes local data to create safer, smarter, and more sustainable communities.

Aligning Edge AI with Strategic Goals

To reap the full benefits of edge AI, organizations must start with clear objectives:

1. Define the critical problems edge AI can solve in your organization – whether it's improving operational efficiency, enhancing customer experiences or mitigating risk.

2. Refine and prioritize the list of critical problems. Determine the necessary requirements and allocate resources and budget to ensure the feasibility and impact of your initiatives. This step lays the groundwork for focused and actionable strategies, aligning efforts with both organizational needs and opportunities.

3. Invest in building a scalable foundation. This includes deploying platforms optimized for edge environments and designing systems that can adapt to future demands.

The unique ability of edge AI deployments to process data locally, reduce latency, and enhance security make them a vital asset for organizations ready to innovate and thrive. By aligning edge AI with the biggest opportunities, organizations can unlock real-time intelligence and redefine what's possible in their industries.
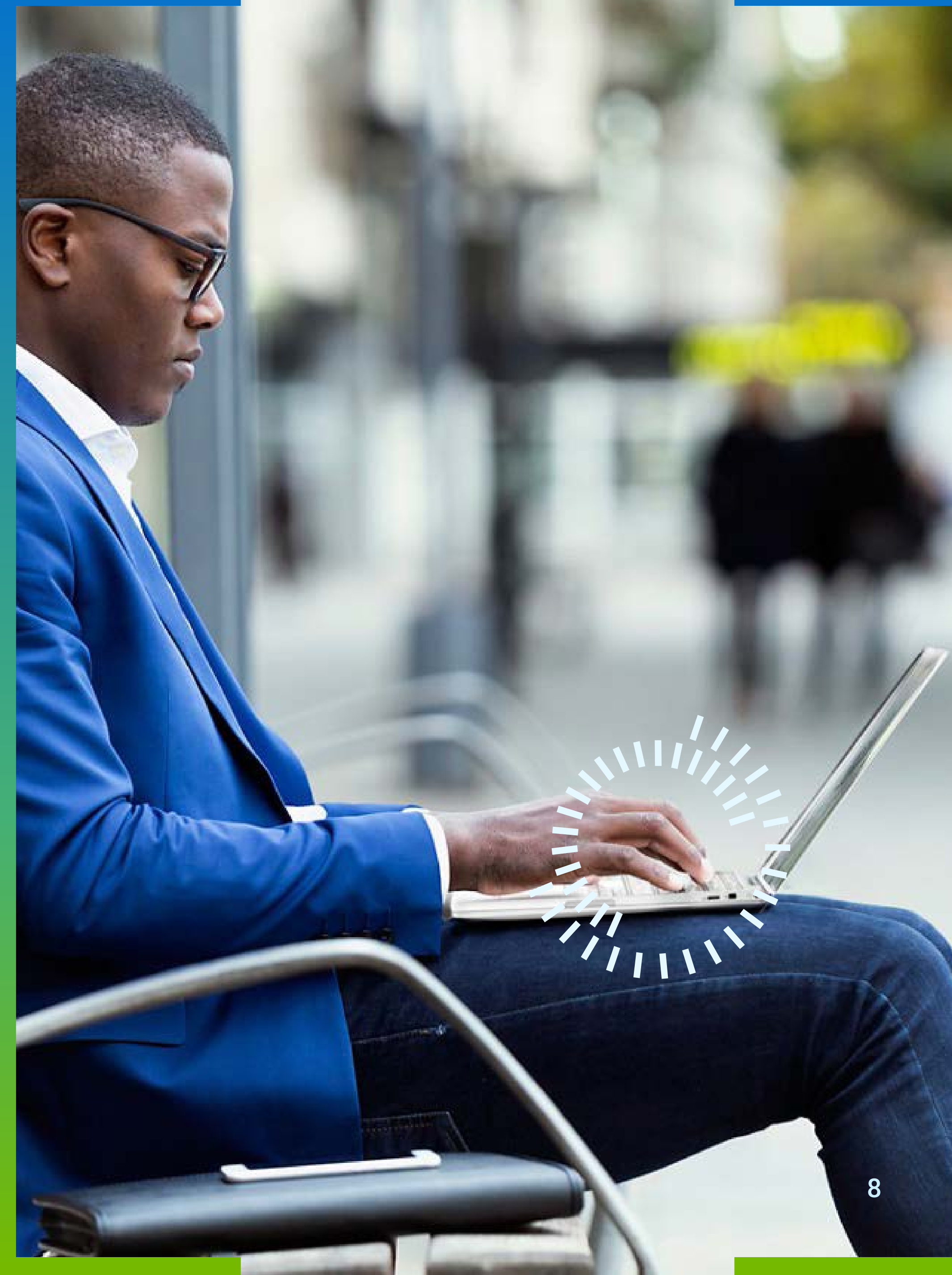
The Challenges of Edge AI

While edge AI provides immense opportunities, organizations must first overcome a range of complexities to successfully bring intelligence closer to the point of action. These challenges align directly with the key requirements of speed, scale, and security, while revealing the intricacies of edge AI environments.

Managing Distributed Systems

Edge AI operates across a web of devices and environments, often spread across diverse geographies. Unlike core data centers or public clouds, these distributed systems require seamless coordination to maintain efficiency and consistency. Managing operations at each edge node—deployed throughout factories, hospitals, or remote sites—demands robust orchestration tools capable of handling diverse use cases and infrastructure without compromising performance.

Real-Time Functionality

The ability to make split-second decisions is a hallmark of edge AI, but latency remains a critical challenge. Use cases like predictive maintenance, smart public safety systems, and personalized retail experiences can require consistent availability and immediate processing of large quantities of edge data. Centralized IT infrastructures often fall short in providing the speed necessary for real-time AI inferencing at the edge, hindering operational outcomes.

Scaling Across Diverse Environments

Adaptability is vital due to the unique needs of each edge AI deployment site, from factories and retail outlets to busy traffic intersections. Scaling solutions across these varying environments can be challenging due to differences in hardware, connectivity, and operational workflows. Without a strategy for consistent and scalable integration, organizations risk inefficiencies and bottlenecks as they expand their systems.

Addressing Security Vulnerabilities

The distributed nature of the edge makes it prone to security risks. Processing data across numerous edge locations significantly increases the surface area for malicious attacks, while transmitting data between edge and a core data center or public cloud environment can introduce further vulnerabilities. Maintaining the integrity of edge devices and processing data using zero-trust security principles is non-negotiable and can be resource-intensive.

Overcoming Edge AI Challenges

Successfully addressing these challenges requires organizations to deploy a thoughtful, strategic approach, with tools and frameworks designed specifically for each unique edge environment.
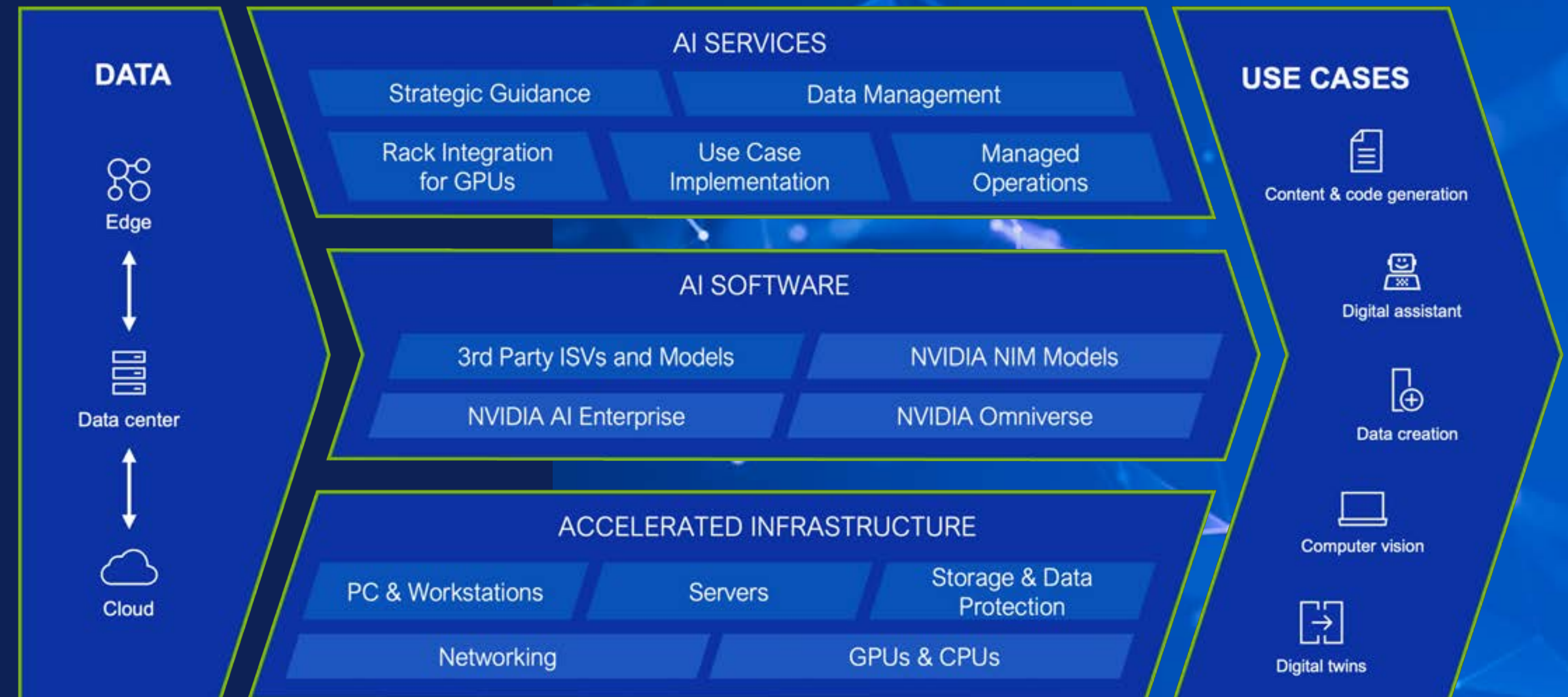
The Dell AI Factory With NVIDIA

The Industry's First End-to-End Enterprise AI Solution [4]

The Dell AI Factory with NVIDIA accelerates AI adoption by delivering combined Dell and NVIDIA® capabilities to power your AI use cases, integrate your data and workflows, and achieve repeatable, scalable outcomes. Used as part of your AI journey, you'll experience:

- ☑ Faster time to value
- ☑ Optimized productivity
- ☑ Transformative insights
- ☑ Outcomes you can trust

DATA
- Edge
- Data center
- Cloud

AI SERVICES
- Strategic Guidance
- Data Management
- Rack Integration for GPUs
- Use Case Implementation
- Managed Operations

AI SOFTWARE
- 3rd Party ISVs and Models
- NVIDIA NIM Models
- NVIDIA AI Enterprise
- NVIDIA Omniverse

ACCELERATED INFRASTRUCTURE
- PC & Workstations
- Servers
- Storage & Data Protection
- Networking
- GPUs & CPUs

USE CASES
- Content & code generation
- Digital assistant
- Data creation
- Computer vision
- Digital twins

[4] Based on Dell analysis, July 2024.

# A Strategic Framework to Make AI Use Cases Real

The Dell AI Factory with NVIDIA transforms data into insights at unparalleled scale and speed. And while much of today's AI-related activities such as modeling and fine-tuning occurs in centralized locations, the real breakthrough happens when you extend the AI Factory to the edge. By dispersing intelligence closer to where data is generated, you can eliminate latency while reducing the cost and complexity of moving vast datasets to and from centralized systems.

Edge devices powered by advanced AI models process these vast streams of real-world data, unlocking insights in milliseconds. From detecting quality defects on production lines to ensuring safe navigation for autonomous vehicles, inferencing at the edge enables immediate action – grounded in the unparalleled intelligence of the Dell AI Factory with NVIDIA framework.
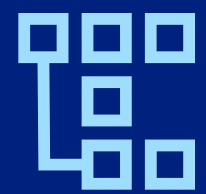
Data

Data is the fuel for the AI Factory. With data increasingly generated at the edge, you need an expert AI services team to help you prepare data for seamless ingestion into the AI platform, ensuring valid and impactful outputs.
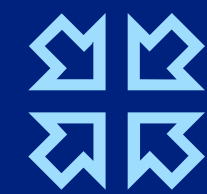
Infrastructure

From servers and storage to networking and PCs, you need infrastructure optimized for edge conditions. Defeat temperature swings, dust and humidity with infrastructure you can count on to deliver high availability and low latency in critical environments – from battlefields to farm fields.

Ecosystem

Harness the power of an ecosystem of technology partnerships, enabling diverse AI applications, models, and frameworks for maximum flexibility. This will ensure you have the capabilities, solutions and tools to innovate faster, cut costs and increase value from data.
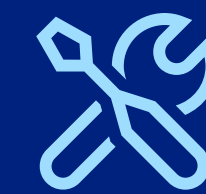
Services

Simplify and accelerate your edge AI journey by combining your unique expertise with a comprehensive suite of edge and AI services, giving you the confidence that your edge AI deployments deliver value more quickly and are optimized, secure and resilient.

Leverage Edge AI:
Digital Twins and Computer Vision

—

Building on the transformational power of the Dell AI Factory with NVIDIA, edge AI applications like Digital Twins and Computer Vision demonstrate how advanced technology can deliver immediate, actionable insights where they matter most.

Digital Twins

Digital twins are virtual representations of real-world entities and processes, synchronized at a specified frequency and fidelity. You can obtain valuable insights by making a digital copy of a product or process to test, analyze, and predict outcomes. The digital twin sector is expected to reach over $140 billion by 2031.[5]

Deliver a wide range of business benefits:

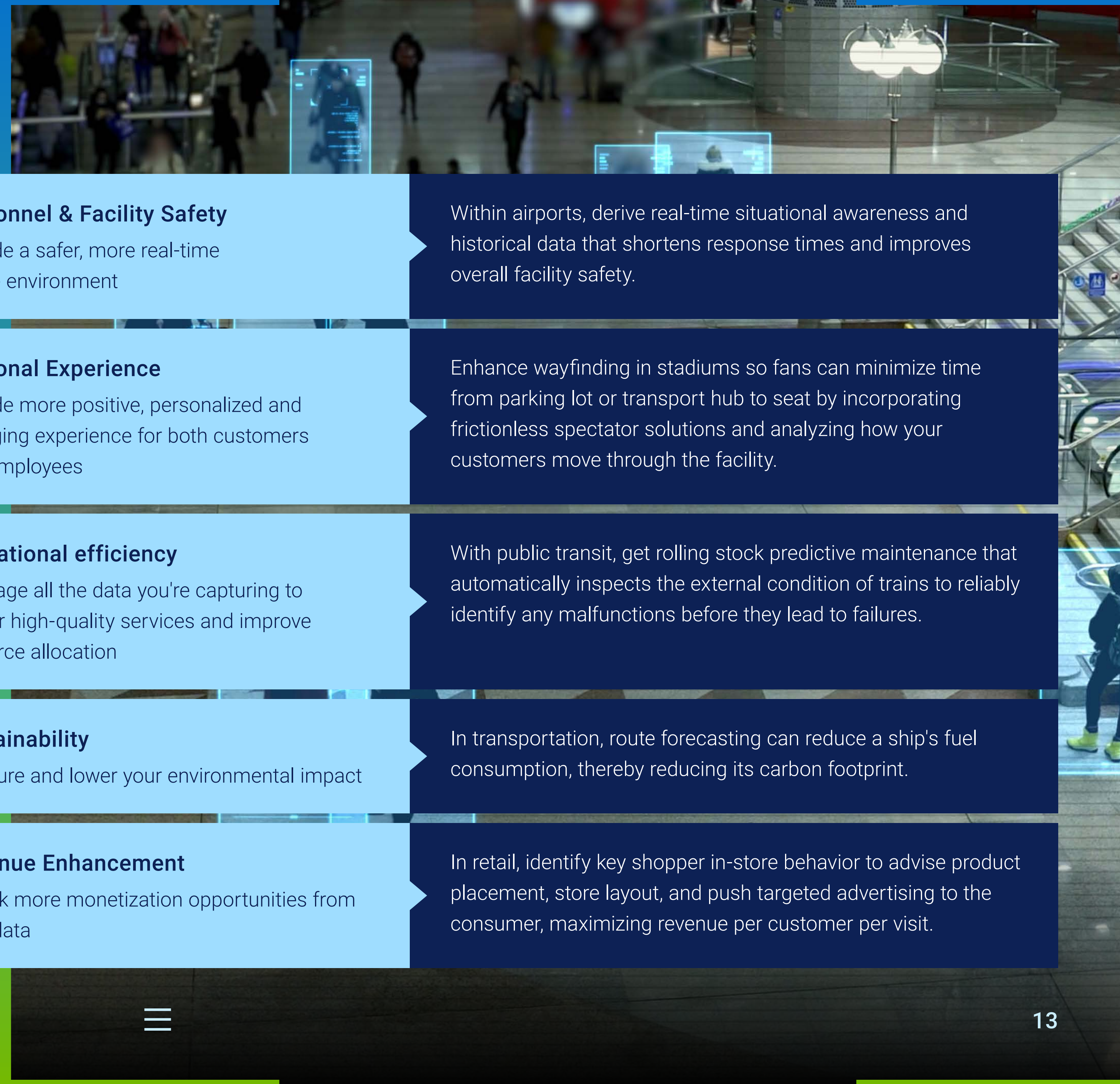|		
⚠ Mitigate risks	Improve customer service	Optimize operations
Increase design quality	Enhance security	Reduce time-to-market

[5] Research and Markets, "Digital Twin Market Size and Forecast 2021-2031, Global and Regional Share, Trend, and Growth Opportunity Analysis Report," June 2024

Computer Vision

Organizations from every industry are in a digital race to turn data into business outcomes faster. Computer vision plays a central role in that effort with its ability to connect video and other data at disparate edge and centralized locations to derive actionable insights.

According to STL Partners, computer vision will account for more than half of the edge AI addressable market by **2030**.[6]

Personnel & Facility Safety

Provide a safer, more real-time aware environment

Within airports, derive real-time situational awareness and historical data that shortens response times and improves overall facility safety.

Personal Experience

Provide more positive, personalized and engaging experience for both customers and employees

Enhance wayfinding in stadiums so fans can minimize time from parking lot or transport hub to seat by incorporating frictionless spectator solutions and analyzing how your customers move through the facility.

Operational efficiency

Leverage all the data you're capturing to deliver high-quality services and improve resource allocation

With public transit, get rolling stock predictive maintenance that automatically inspects the external condition of trains to reliably identify any malfunctions before they lead to failures.

Sustainability

Measure and lower your environmental impact

In transportation, route forecasting can reduce a ship's fuel consumption, thereby reducing its carbon footprint.

Revenue Enhancement

Unlock more monetization opportunities from your data

In retail, identify key shopper in-store behavior to advise product placement, store layout, and push targeted advertising to the consumer, maximizing revenue per customer per visit.

Bring the Dell AI Factory to the Edge with Dell NativeEdge

Dell NativeEdge brings the power of Dell AI Factory with NVIDIA to the edge by enabling organizations to securely scale their infrastructure and orchestrate AI applications across any location. Support for virtualized and containerized environments is seamless, while NativeEdge Blueprints automate the deployment of frameworks and applications for faster, more efficient AI innovations. Dell NativeEdge can also be configured with high-performance 5G connectivity that meets the rigorous security, reliability, and performance demands of even the most comprehensive edge deployments.

Save time and reduce costs with NativeEdge zero-touch deployment and automated application orchestration.

Centralize management of diverse devices and edge AI applications, and **free up time** with automated operations.

Stay ahead of competition by rapidly and easily deploying your choice of new AI use cases powered by Dell and NVIDIA.

Benefits of Dell NativeEdge

The world's most **adaptable and open** edge operations ecosystem [7]

Up to **68% time savings** by automating edge application orchestration [8]

Less than 1 minute to deploy infrastructure and applications [8]

Dell NativeEdge: Accelerate AI innovation at the edge

Zero-touch onboarding

Zero-trust security

Multicloud orchestration

Edge operations software solution

Infrastructure management   Application orchestration

NativeEdge OS   Containers   Virtual machines

Why Dell NativeEdge and NVIDIA

Dell NativeEdge is the first edge orchestration solution that automates the delivery of NVIDIA AI Enterprise software, bringing NVIDIA frameworks for video analytics, speech and translation, and optimized inferencing to your edge devices.

This capability is powered by NativeEdge Blueprints, which act like a recipe, detailing the ingredients and steps for automated deployment. This includes application settings, infrastructure resources, network configurations, and custom workflows.

This tight integration between Dell NativeEdge and NVIDIA delivers:

Faster, easier deployment of AI models and inferencing solutions to distributed edge locations.	Zero-touch deployment of edge AI apps across infrastructure based on zero-trust security principles to meet even the most rigorous standards.	Ease of Edge AI solution lifecycle management with scalability to support the unique requirements of any workload.

Infrastructure is the Foundation of the Dell AI Factory with NVIDIA

The Dell AI Factory with NVIDIA brings together Dell AI Infrastructure for powerful computing and networking, enhanced by NVIDIA acceleration, NVIDIA AI Enterprise software, and Dell Professional Services, to form a seamless, all-in-one solution for businesses ready to unlock the power of AI. Our edge-optimized platforms for AI provide a wide range of capabilities, such as:

- Ruggedized platforms
- GPU-enabled systems
- Small form-factor products
- Long life systems
- Industry certifications
- OEM-Ready solutions (de-branded, re-brand ready)
- Customization capabilities

PowerEdge Servers
Fast-track your Edge AI goals using PowerEdge servers with superior acceleration, diverse GPU options and rugged, small footprint and industry certified platforms.

Dell Laptops and Workstations
Allow AI processing locally on the device. Our broad portfolio of AI workstations and AI PCs provide the necessary hardware and software infrastructure to enable AI inferencing at the edge, empowering organizations to leverage the power of AI in real-time, even in resource-constrained environments.

Edge Gateways
Compact in design, Edge Gateways enable you to collect, consolidate, and extract value from vast amounts of edge-generated data.

Storage and Data Protection
Unlock the value of edge data and secure AI workloads against data loss and cyber-threats with modern, simple, and resilient storage and data protection solutions.

Hyperconverged Infrastructure (HCI)
Benefit from the breadth of the Dell Technologies HCI portfolio that allows for choice based on your desired outcomes.

Networking
Enable simplified design, management and monitoring of powerful Ethernet fabrics to handle modern workloads like Generative AI and Edge AI inferencing.

Data Management
Empower your customers to access data across edge, core and multicloud to power analytics workloads for faster model tuning and business insights.

Enabling Edge AI Innovation with Advanced Server Hardware

As organizations navigate an increasingly digital-first world, the demand for secure, high-performance infrastructure has become a top priority. Dell PowerEdge XR servers, powered by NVIDIA GPUs, are engineered to meet these demands head-on, delivering powerful, scalable solutions for AI applications at the edge. With real-time data analysis capabilities, these servers ensure low latency and high performance beyond the traditional data center, even in space-constrained or challenging environments.

Built to withstand the toughest environments:

- Extreme heat and cold
- Dust
- Shock and vibration of factory floors
- Construction sites
- Mobile command centers
- Other extreme environments

Unlock a Data-Powered Future with Edge AI

With trends like real-time personalization, urban optimization, and adaptive AI ecosystems on the rise, businesses that adopt edge strategies today will define the standards for tomorrow. Integrating edge AI today drives new efficiencies while fostering sustainable operations through reduced waste and optimized natural resource use.

Take ownership of this future by positioning your organization as a leader in edge-driven AI innovation.

Build your AI-Driven edge today

Finding the right partners to guide your AI transformation is key. Dell Technologies and NVIDIA provide the technology and tools needed to help businesses succeed at the edge, from ruggedized servers powering field applications to NativeEdge simplifying deployments at scale.

Together with Dell and NVIDIA, you can turn potential into action by unlocking the power of your data where it lives, ushering in a new era of success for your organization.

87%
of the Fortune 100 use Dell Technologies edge solutions [9]

Dell Technologies is the **#1** global leader in edge storage and edge servers[8]

[9] Dell Technologies internal analysis of US Fortune 100, February 2025

[10] IDC Worldwide Quarterly Enterprise Infrastructure Tracker: Buyer and Cloud Deployment, Q4 2024, March 2025. Based on revenue. Edge storage and edge servers refer to the heavy edge category, which IDC defines as the physical equipment that serves as the foundational infrastructure for edge.

Take the Next Step

The **Dell Accelerator Workshop** is a great first step for institutions looking to begin their AI and edge journey. This half-day program focuses on the activities required to achieve your desired end state, concluding with next steps to further advance your business and IT strategies.

Your team will work with Dell experts to develop a point of view on important GenAI questions and create a vision for your future state. Utilizing our "AS-IS" / "TO-BE" methodology, we'll conduct interviews and review your existing environment to identify challenges, opportunities and drive consensus for GenAI, synthesized in an Executive Overview.

Interested in validating the capabilities of NativeEdge in your edge environment? Contact your Dell sales representative to discuss the possibility of deploying an onsite proof-of-concept experience for Dell NativeEdge.
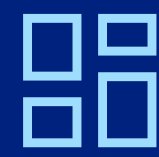
Accelerate and simplify your AI journey with Professional Services

Develop a Generative AI strategy and roadmap tailored to your organization

Prepare your data for Generative AI integration, inferencing, and model customization

Build your Generative AI operations with training and infrastructure management expertise

Assess current state

Identify challenges in the environment

Define goals and objectives

Establish vision for ideal future state

Develop an executive roadmap

Define expected results and benefits

Harness the power of data generated at the edge to meet the evolving and rigorous requirements of your industry with the Dell AI Factory with NVIDIA and Dell NativeEdge.

Dell **AI Factory**

WITH NVIDIA

DELLTechnologies | NVIDIA

Learn more about the Dell AI Factory with NVIDIA ›

Learn more about Dell NativeEdge ›