

HPE ProLiant Gen11 servers with Clarifai for manufacturing (edge and data center solutions)

Introduction

The future of manufacturing, powered by artificial intelligence (AI) at the edge and in the data center, is set to dramatically reshape the industry. Depending on the need, AI at the edge is expected to see rapid adoption, with IDC predicting that by 2025, 50% of new IT infrastructure deployed by manufacturers will be at the edge.¹ This is driven by the need for real-time data processing and decision-making in environments such as factories and industrial plants, where latency and downtime can lead to significant operational issues and losses. Meanwhile, data generation in the manufacturing sector is projected to skyrocket; McKinsey estimates that by 2025 IoT will have a potential total economic impact of as much as \$11.1 trillion per year.² Processing and managing this data volume requires a mix of edge, on-premises, and hybrid solutions to optimize both speed and scale.

¹ ["IDC FutureScape: Worldwide Cloud 2024 Predictions,"](#) IDC, 2023.

² ["How can we recognize the real power of the Internet of Things?"](#) McKinsey & Company, 2017.

AI use cases such as predictive maintenance and quality control inspection (QCI) are already delivering measurable results. The median unplanned downtime costs for unplanned downtime exceeded \$100,000/hr.³ Similarly, McKinsey reported that AI based use-cases in manufacturing saw incredible results, including a two to three times increase in productivity, a 50 percent improvement in service levels, a 99 percent reduction in defects, and a 30 percent decrease in energy consumption.⁴ These applications highlight AI's potential to unlock significant value in manufacturing, leading to both costs and savings and increased productivity across the board.

HPE ProLiant Compute and Clarifai

HPE ProLiant Compute is a family of scalable and high-performance servers designed for diverse workloads including AI, data analytics, virtualization, and edge computing. These servers offer flexibility with various configurations (tower or rack mounted) and support hybrid cloud environments. Key features include scalability, security through silicon root of trust from HPE, and the latest processors such as Intel® Xeon® Scalable processors and AMD EPYC. HPE ProLiant servers are ideal for businesses needing reliable, secure, and powerful computing solutions for data centers, enterprise applications, and AI workloads.

Clarifai, a leading AI company specializing in computer vision and machine learning (ML), was established in 2013. Clarifai is available in cloud, on-premises, or hybrid environments and has been used to build more than 1.5 million AI models with more than 400,000 users in 170 countries. Clarifai has been recognized as a strong performer and a leader in Forrester 2024 evaluations for computer vision and AI platforms.⁵ Their advanced technology, including capabilities for image and video recognition, is particularly effective for manufacturing use cases such as quality control and predictive maintenance.

Industry challenges and opportunities

— Common challenges faced in manufacturing

- Equipment downtime: Unplanned machinery breakdowns lead to costly downtime and reduced productivity
- Quality control (QC): Maintaining consistent product quality while reducing defects remains a critical challenge in high-volume production
- Cost control: Rising raw material costs and energy expenses pressure manufacturers to maintain profitability
- Supply chain disruptions: Manufacturing operations are often impacted by unpredictable supply chain delays and shortages

— Opportunities with advanced HPE ProLiant servers and Clarifai AI solutions

- Predictive maintenance: HPE ProLiant servers along with Clarifai platform enable real-time monitoring of equipment, using AI to predict failures and help minimize unplanned downtime.
- Automated QC: HPE AI solutions can automate defect detection, helping ensure consistent product quality and reducing waste in high-volume production lines.
- Cost efficiency through AI: AI-driven optimization on HPE servers can reduce energy consumption and material waste, improving cost control.
- Supply chain optimization (SCO): AI-powered analytics can enhance supply chain visibility, helping manufacturers predict disruptions and adjust proactively.

³ ["Predictive maintenance market: 5 highlights for 2024 and beyond,"](#) IoT Analytics, 2023.

⁴ ["How manufacturing's Lighthouses are capturing the full value of AI,"](#) McKinsey & Company, 2024.

⁵ ["Clarifai named a Leader in The Forrester Wave™: Computer Vision Tools, Q1 2024,"](#) Clarifai, 2024.

High-level overview of the solution

Figure 1 illustrates that in an edge AI computer vision scenario for manufacturing, cameras capture real-time video or images of the production process, and the data is processed locally by edge servers or devices, such as the HPE ProLiant DL 145 Gen11 server. AI models analyze the data for tasks such as QCI or predictive maintenance, detecting defects, anomalies, or wear in machinery. The local processing reduces latency and allows immediate action on insights while avoiding reliance on cloud connectivity. Components include edge servers, AI models, cameras, sensors, and industrial control system.

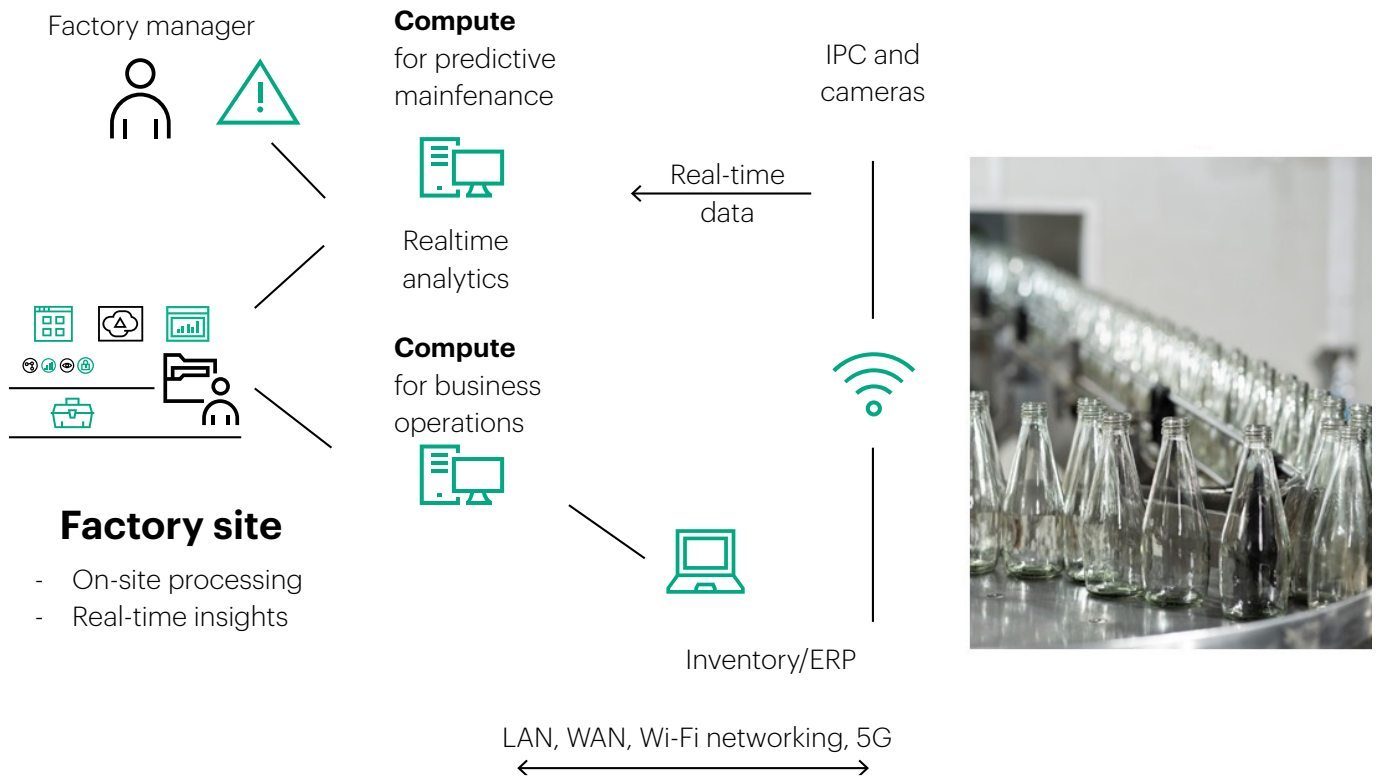


Figure 1. AI at edge for manufacturing

Figure 2 shows the Clarifai's architecture that is modular and scalable, built around an AI engine with pretrained and custom model capabilities. Data flows through APIs into a pipeline where it's labeled and stored, enabling model training and deployment. The control plane manages operations through API while the compute plane handles model training and inference. Web apps provide the presentation layer for users to interact with data and models. The platform integrates with cloud or on-premises environments, ensuring secure and efficient processing. It includes robust logging and monitoring for performance tracking and troubleshooting.

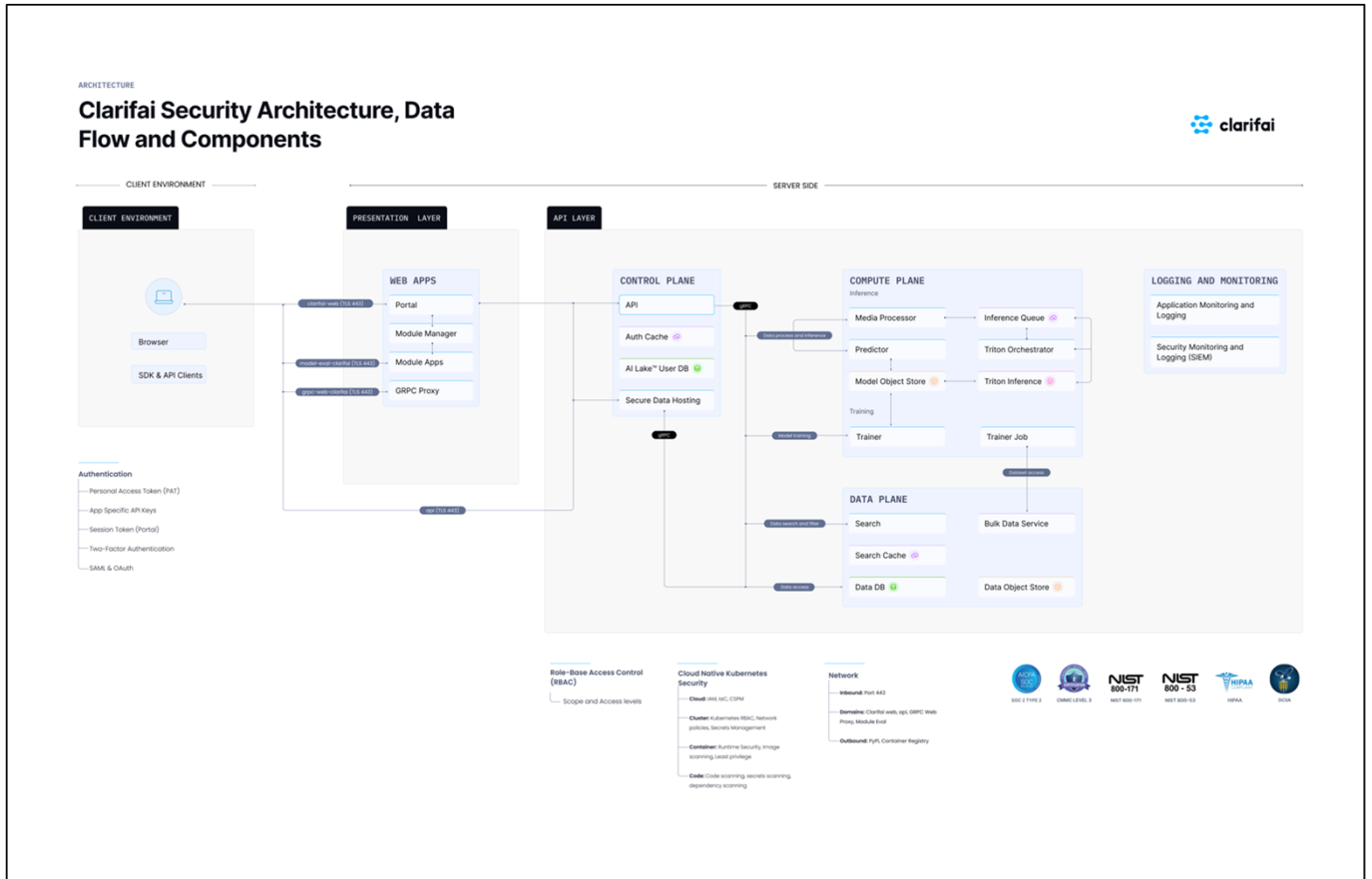


Figure 2. Clarifai data architecture and solution overview

Key solution components

Clarifai: Intelligence layer

- Clarifai is an AI platform specializing in computer vision, natural language processing, and deep learning.
- It provides tools for building, training, and deploying ML models, especially in image, video, and text recognition.
- It serves industries such as manufacturing, retail, defense, healthcare, media, and logistics for tasks such as visual search, content moderation, and predictive analytics.

Key applications

- Image and video recognition: Automatic identification and tagging of objects, faces, and scenes in visual data
- Content summary: Summarization of key manufacturing artifacts including manuals, reports, and design specification documents
- Visual search: Search facilitation using images rather than text, improving product discovery

- Predictive maintenance: Detection of potential failures in machinery through image-based inspections
- Document analysis: Extraction of text, forms, and relevant information from documents using Optical Character Recognition (OCR)

Key features

- Pretrained models: Offers a wide range of ready-to-use AI models for object detection, face recognition, and text processing
- Custom model training: Allows users to fine-tune models using their own data
- Edge deployment: Allows models to be deployed on edge devices for low-latency AI processing
- Multimodal AI: Supports combining different types of data (for example, text and images) for more complex applications
- API integration: Facilitates easy integration into existing systems through RESTful APIs and software development kits (SDKs)

Clarifai platform architecture / Computer Vision Architecture

- Cloud and edge AI: Supports both cloud-based model deployment and edge computing for real-time analysis
- Unified data platform: Provides a single platform for managing and annotating data, training models, and deploying them
- Scalable architecture: Enables handling of large-scale data and model training, leveraging GPUs and parallel processing
- Data annotation tools: Offers integrated tools for annotating and managing datasets for model training
- Prebuilt models and model zoo: Offers a library of pretrained models across various domains (for example, object detection and language processing)

Key benefits

- Speed and accuracy: Delivers high-speed image and video recognition with accurate results
- Customizability: Allows fine-tuning of models to meet specific business needs and industries
- Scalability: Enables processing of massive datasets, suitable for enterprises with large data requirements
- Easy integration: Facilitates seamless integration with existing workflows through APIs, enabling faster implementation
- Reduced development time: Shortens the time to market for AI applications with prebuilt models and training tools.

Key differentiation

- Pretrained and custom models: A combination of pretrained models and the ability to create custom models offers flexibility for various use cases.
- Edge deployment: A support for deploying AI models at the edge enables low latency processing critical for real-time applications.
- Multimodal AI: The ability to combine data types (image, video, text) into a unified model allows more complex and comprehensive solutions.
- Full-stack AI platform: Clarifai provides a complete end-to-end AI lifecycle solution, from data labeling to deployment, unlike some competitors who focus on specific aspects.
- Wide industry coverage: Clarifai serves a diverse range of industries, making its platform highly versatile and adaptable.

- Memory: 32 HPE DDR5 Smart Memory RDIMMs
8 channels per CPU, up to 4800 MT/s
- GPU: Up to eight NVIDIA® L4 or three L40S Tensor Core GPUs
- Storage: 8 SFF supported
- Storage controller: HPE MR408i-o Gen11 x8 Lanes
4GB Cache OCP SPDM Storage Controller
- Management: HPE iLO 6 and HPE GreenLake for
Compute Ops Management
- Processor: 5th Gen Intel Xeon Scalable processors

For complete spec information, see the [QuickSpecs of HPE ProLiant DL380 Gen11 Server](#)



Figure 3. HPE ProLiant DL380 Gen11 Server

Optimized performance for AI

The HPE ProLiant servers are specifically designed to meet the rigorous demands of AI workloads, offering unparalleled performance, scalability, and reliability. These servers are equipped with the latest advancements in processing power, including support for high performance CPUs and GPUs, which are essential for handling complex AI algorithms and large datasets. Additionally, the Gen11 servers boast advanced memory capabilities and high-speed interconnects, helping ensure rapid data processing and real-time analytics. By integrating these cutting-edge technologies, HPE ProLiant Gen11 servers deliver the computational horsepower needed to accelerate AI applications and enhance decision-making processes across various industries.

Furthermore, HPE ProLiant Gen11 servers are built with a focus on seamless integration and flexibility making them an ideal choice for AI-centric environments. The servers support a broad range of AI frameworks and software, helping ensure compatibility with industry-standard tools.

Securely access, monitor, and manage servers

Delivering a modern compute management experience through HPE GreenLake cloud, HPE GreenLake for Compute Ops Management transforms your entire environment to drive better efficiency of server deployments, provides secure connections from edge and data center to cloud, and automates tasks for a low-touch experience. Now, you can remotely manage your server estates or create new managed services for your customers and lines of business.

With HPE GreenLake for Compute Ops Management, you can:

- **Do more:** Automated operations replace manual tasks and let you deploy and manage your compute environment with ease.
- **Move faster:** With efficient server updates, your teams get back to their evenings and weekends.
- **Set and protect:** Faster, continuous patching, policies, and secure connections help you reduce security risk.

Trusted security by design

Protect your infrastructure, workloads, and data from threats to hardware and risks from third-party software with a proven, zero trust approach to security. The HPE iLO ASIC acts as a silicon root of trust -- making it virtually impossible to insert any malware, virus, or compromised code that would corrupt the server boot process.

Securing servers in today's expansive IT landscape is a multifaceted challenge. HPE ProLiant servers meet the challenge by embedding security features at every level.

- **HPE trusted supply chain security** giving you confidence that your hardware is secure at the source
- **Secure bootup and hardware-level security measures** with silicon root of trust from HPE, security built into the very core of technology
- **Securely decommission servers** and protect your data by using the one-button secure erase or system erase and reset features
- **Secure, authenticated, and encrypted connection** from your servers to HPE GreenLake for Compute Ops Management service helps ensure you can remotely access and maintain control

Business applications and use cases

Predictive maintenance

- **Impact:** Predictive maintenance can result in a 5%–15% reduction in facility downtime
- **Cost savings:** Increase labor productivity by 5%–20%⁶
- **Strategic advantage:** Increased asset utilization and extended equipment life, positioning companies for long-term operational efficiency and reliability
- **Business outcome:** Improved uptime and equipment availability lead to higher overall production output, resulting in a 3%–5% reduction in equipment costs⁷

^{6,7} ["Predictive maintenance Deloitte's approach,"](#) Deloitte, 2022.

QCI

- Impact: 90% reduction in defects with AI-powered computer vision systems⁸
- Cost savings: Lowered defect-related losses and product recalls, resulting in an increase in profitability
- Strategic advantage: Improved product quality and brand reputation, gaining a competitive edge in markets that demand high precision and reliability
- Business outcome: Enhanced customer satisfaction and reduced warranty claims, leading to stronger market positioning and boost in customer retention

SCO

- Impact: AI-driven demand forecasting improves inventory management, reducing holding costs
- Cost savings: Reduced stockouts, excess inventory, and expedited shipping costs, translating to an increase in revenue
- Strategic advantage: Enhanced supply chain agility, enabling faster response to market fluctuations and securing better supplier relationships
- Business outcome: Optimized inventory management leads to improved cash flow, allowing businesses to reinvest in growth opportunities and innovation

Production process automation

- Impact: Automated workflows through AI reduce production cycle times
- Cost savings: Lower labor costs and reduced production delays, saving millions annually
- Strategic advantage: Increased production capacity without additional capital expenditure, allowing businesses to quickly scale operations and meet market demands
- Business outcome: Accelerated time to market (TTM) for new products, enhancing competitiveness and leading to increase in market share in certain industries

Energy efficiency optimization

- Impact: AI-driven energy management systems reduce energy consumption
- Cost savings: Significant reductions in energy costs, often cutting operational expenses by millions per year
- Strategic advantage: Improved sustainability, meeting regulatory and market demands for green manufacturing while also lowering the carbon footprint and improving brand perception
- Business outcome: Increased sustainability efforts improve relationships with stakeholders and investors while also contributing to increase in long-term profitability through cost efficiencies and brand loyalty



⁸ [AI in Manufacturing: Benefits, Use Cases, and Examples](#), imaginovation.net, 2023.



Next steps

HPE ProLiant Gen11 servers are built with a focus on seamless integration and flexibility, making them an ideal choice for AI-centric environments. The servers support a broad range of AI frameworks and software, helping ensure compatibility with industry-standard tools. The HPE ProLiant servers also come with robust security features, including silicon root of trust from HPE and advanced encryption, to safeguard sensitive data and maintain compliance with industry regulations. With our comprehensive management capabilities, organizations can easily deploy, monitor, and manage their AI workloads, helping ensure optimal performance and minimal downtime. In summary, HPE ProLiant Gen11 servers provide a powerful, secure, and adaptable foundation for businesses looking to harness the full potential of AI.

Learn more at

[HPE.com/ProLiant](https://hpe.com/ProLiant)

[HPE.com/ProLiant/solutions](https://hpe.com/ProLiant/solutions)

Visit [HPE.com](https://hpe.com)

[Chat now](#)

© Copyright 2025 Hewlett Packard Enterprise Development LP. The information contained herein is subject to change without notice. The only warranties Hewlett Packard Enterprise products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. Hewlett Packard Enterprise shall not be liable for technical or editorial errors or omissions contained herein.

AMD is a trademark of Advanced Micro Devices, Inc. Intel Xeon is the trademark of Intel Corporation or its subsidiaries in the U.S. and/or other countries. NVIDIA is a trademark and/or registered trademark of NVIDIA Corporation in the U.S. and other countries. All third-party marks are property of their respective owners.

a00143230ENW

HEWLETT PACKARD ENTERPRISE

hpe.com

