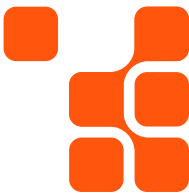


CLUSTERA AI INFERENCE



Run Production AI Wherever Your Data Lives



As AI agents and applications move into production, organizations need a secure, scalable and efficient way to deploy AI across their enterprise data landscape. Cludera AI Inference Service delivers NVIDIA-powered inference wherever data resides—on premises, in private clouds, or across public clouds—helping organizations build differentiated AI while maintaining data sovereignty and predictable economics.

Key Features & Benefits

Cludera AI Inference provides the foundation for production AI, enabling organizations to deploy, govern, and scale AI applications and agents across hybrid environments with high performance, data sovereignty, and predictable costs.

Capabilities	What it Enables	Outcome
Flexible Deployment	Run AI wherever data resides across on premises and cloud environments	Differentiated AI, Control, and Data Sovereignty
NVIDIA-Powered Inference	Run LLM and agentic workloads with optimized performance	Faster AI Use cases
Governance & Operations	Secure, monitor, and scale AI workloads	Trusted Production AI
Capacity-Based Infrastructure	Align infrastructure to workload demand	Predictable Costs for Running AI

At a Glance

AI is increasingly embedded in core business operations. Organizations need a secure and scalable way to deploy AI across enterprise data. Competitive advantage increasingly comes from combining foundation models with proprietary data and business context.

Cludera AI Inference is powered by NVIDIA technologies for production LLM applications and agents, enabling organizations to run AI wherever their data resides across hybrid environments.

With Cludera, organizations can maintain data sovereignty, control infrastructure costs with flexible capacity-based pricing, and scale AI adoption with confidence.

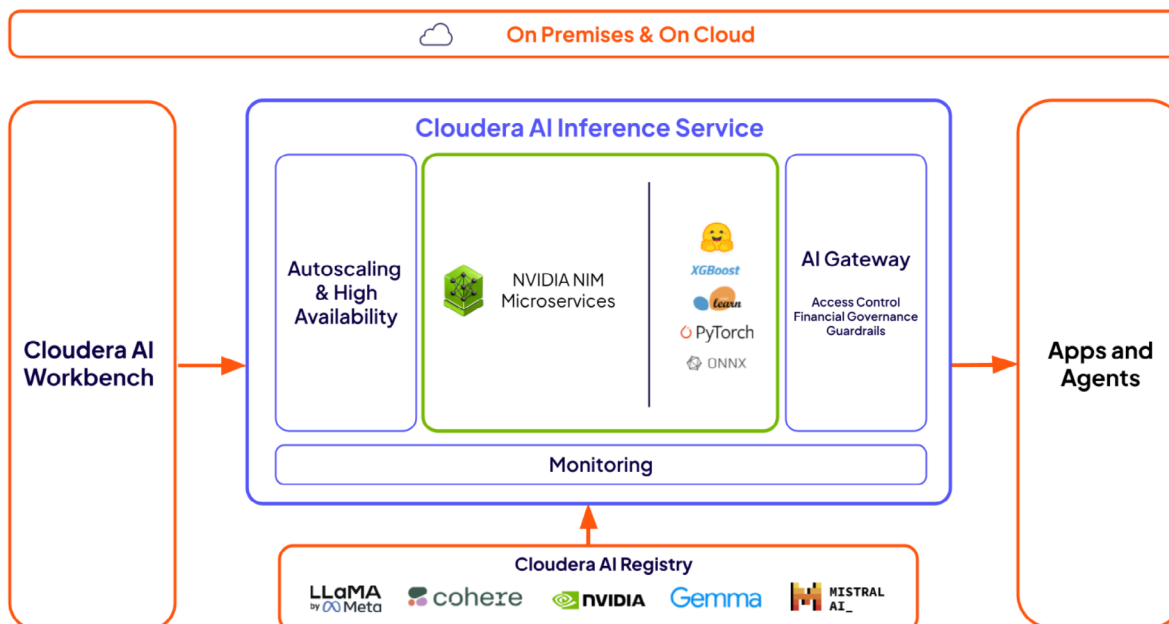


Figure 1. Cloudera AI Inference connects AI development with production deployment, serving models securely and reliably to enterprise applications and agents.

Use Cases and Outcomes

Cloudera AI Inference Service supports deploying AI securely in highly regulated environments across financial services, healthcare, telecommunications, manufacturing, and the public sector.

Use Case	What Cloudera AI Inference Offers	Business Outcomes
Agentic Workflows Claims Processing • Fraud Investigation • IT Incident Resolution	Orchestrate multi-step tasks and manage model interactions	Reliable, scalable AI agents
AI Assistants Developer Copilots • Customer Service • Banking Chatbots • Clinical Assistants	Connect models to enterprise knowledge and generate responses	Faster, more relevant answers
Multimodal AI Medical Imaging Analysis • Contract Review • Manufacturing Quality Check	Run LLMs and specialized models across voice, videos, images, and documents	Process complex data at scale
Real-Time Predictions Credit Risk Scoring • Predictive Maintenance • Demand Forecasting	Serve AI/ML models with low-latency inference at scale	Real-time insights at scale

Cloudera AI Inference brings AI to enterprise data—not the other way around. Organizations can deploy AI securely within their own environments while maintaining control over data, models, and AI operations. Whether supporting AI assistants, agentic workflows, or industry-specific AI applications, Cloudera AI Inference provides the foundation for scalable, governed, and cost-efficient production AI. Explore more [here](#).

CLOUDERA

Cloudera, Inc. | 6220 America Center Dr, 5th Fl, San Jose, CA 95002 USA | cloudera.com

Cloudera is the only hybrid data and AI platform company that large organizations trust to bring AI to their data anywhere it lives. Unlike other providers, Cloudera delivers a consistent cloud experience that converges public clouds, on-prem data centers, and the edge, leveraging a proven open-source foundation. As the pioneer in big data, Cloudera empowers businesses to apply AI and assert control over 100% of their data, in all forms, improving security, governance, and real-time and predictive insights. The world's largest brands across all industries rely on Cloudera to transform decision-making and ultimately boost bottom lines safeguard against threats, and save lives.

To learn more, visit Cloudera.com and follow us on [LinkedIn](#) and [X](#).