

WHITE PAPER

Cloudera and NVIDIA Accelerate AI in the Financial Services Industry

Data, AI, and Analytics, Combined With Performance,
Deliver Outstanding Results

By Stephen Catanzano, Senior Analyst
Enterprise Strategy Group

January 2025

Contents

Abstract.....	3
Financial Services Have a Long List of AI Use Cases	3
Organizations Are Looking for a Multifaceted Data Platform	4
Why the Right Data Platform Is Critical for Financial Firms	5
Hybrid Is a Top Requirement.....	6
Cloudera True Hybrid	7
Cloudera: Data Platform for the Financial Industry.....	7
Cloudera AI With NVIDIA for Fast, Effective AI Development	8
Development Environment	8
Production Environment.....	9
Private AI for Financial Firms	10
Simplified Deployment and Support	10
Conclusion	11

Abstract

In the financial services industry, the ability to harness data, analytics, and artificial intelligence (AI) is critical for maintaining a competitive edge, driving innovation, and unlocking diverse AI use cases. Delivering these capabilities requires a modern data platform capable of handling complex workflows and achieving exceptional performance. In collaboration with NVIDIA, Cloudera provides an advanced solution that meets these demands, enabling financial institutions to accelerate AI initiatives with precision and scale.

This white paper highlights the synergy between Cloudera's AI platform and NVIDIA accelerated computing technologies. Cloudera and NVIDIA empower financial firms to transform data into actionable insights, while driving faster innovation to maintain a competitive advantage in a rapidly evolving landscape.

Financial Services Have a Long List of AI Use Cases

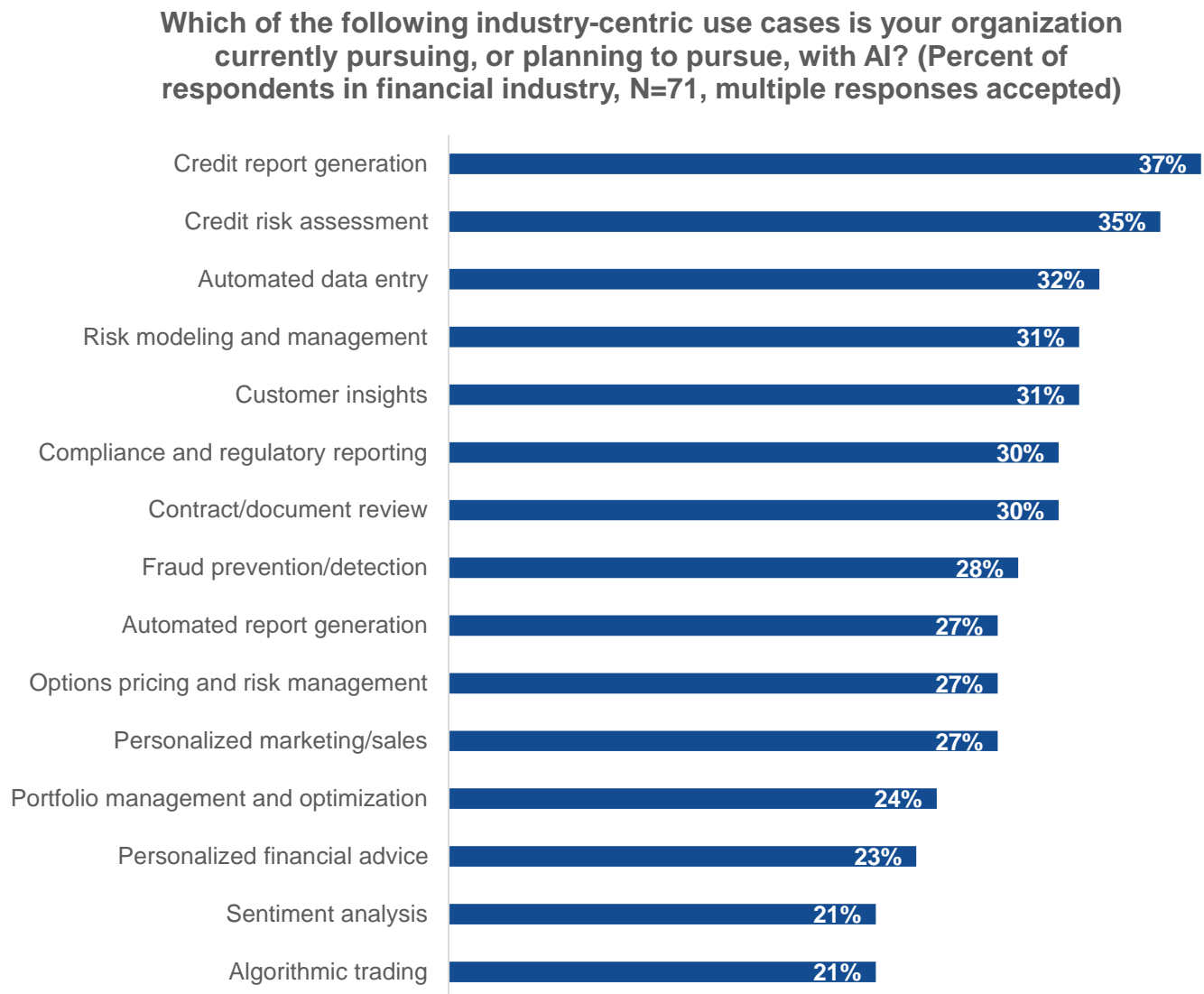
AI is particularly strong in financial services since the industry inherently relies on data-driven decision-making, efficiency, and innovation. AI technologies excel in these areas by offering unparalleled capabilities to process, analyze, and act on large volumes of complex data, enabling financial institutions to maintain a competitive edge.

Informa TechTarget's Enterprise Strategy Group asked financial industry technology decision-makers what industry-specific use cases their organizations are currently pursuing or plan to pursue with AI. The list shown in Figure 1 demonstrates the importance this industry places on using AI capabilities across their organizations.¹ The top areas include:

- **Reducing risk:** Credit report generation, credit risk assessments, risk modeling and management, fraud prevention and detection, options pricing, risk management.
- **Customer experience:** Customer insights, personalized marketing and sales, portfolio management and optimization, personalized financial advice.
- **Automating processes:** Automated data entry, contract document review, sentiment analysis, algorithmic trading.
- **Reporting:** Compliance and regulatory reporting.

Achieving these goals requires a modern data platform designed to manage, govern, and build trusted data at scale to support data, analytics, and AI use cases with very high performance.

¹ Source: Enterprise Strategy Group Complete Survey Results, [The State of Analytics and Business Intelligence Platforms](#), April 2024.

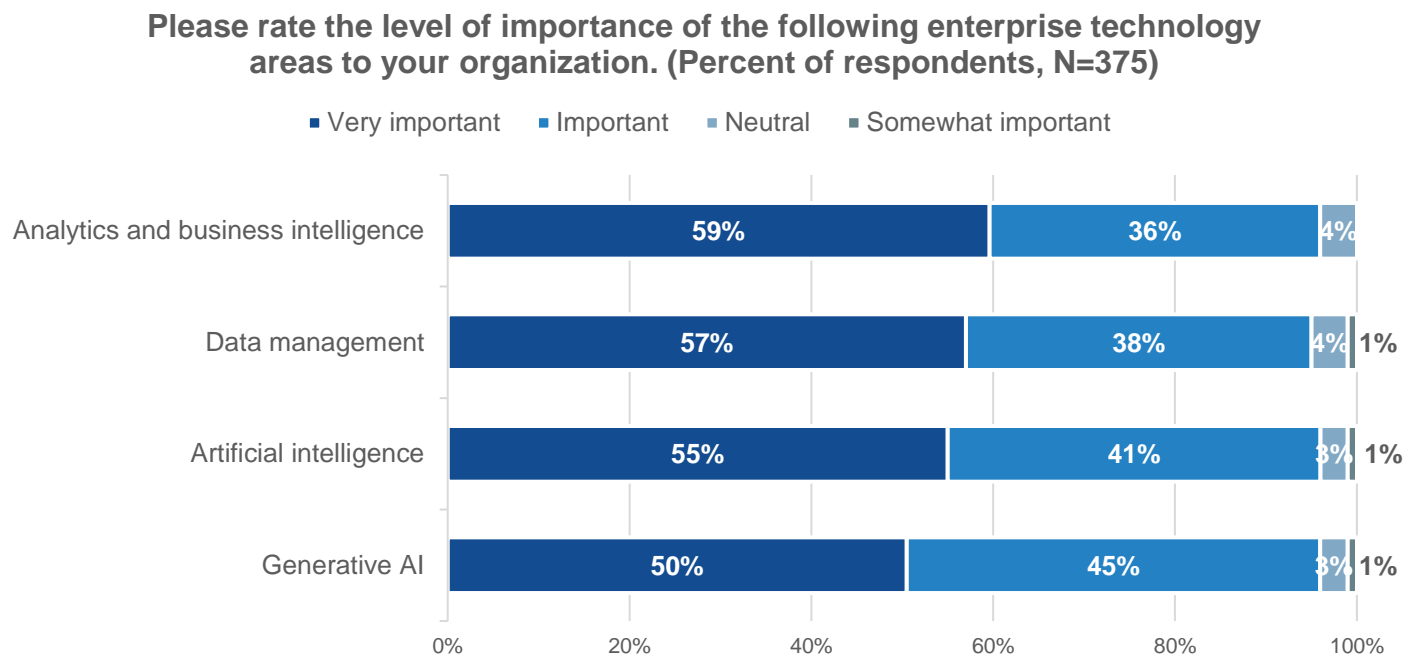
Figure 1. Financial Services Use Cases

Source: Enterprise Strategy Group, a division of Informa TechTarget

Organizations Are Looking for a Multifaceted Data Platform

Enterprise Strategy Group asked IT organizations about the importance of key modern enterprise technology areas. As shown in Figure 2, more than half of respondents felt that analytics and business intelligence (59%), data management (57%), AI (55%), and generative AI (50%) were all very important to their organization.² There is an overwhelming importance placed on technologies that have the ability to drive value from an organization's data.

² Ibid.

Figure 2. Important Technology Considerations

Source: Enterprise Strategy Group, a division of Informa TechTarget

Cloudera's true hybrid platform delivers the enterprise technology that global financial firms demand and effectively brings technology to the data wherever it might reside to address the critical importance of data for financial firms as a central core of their organization.

Why the Right Data Platform Is Critical for Financial Firms

Financial firms have unique challenges driving their need for AI-enabled data platforms to help them operate at scale, remain competitive, and continuously innovate. Their data-driven nature, data sensitivity, scale, and high levels of customer engagement also require strong governance and resiliency. Financial firms' strong use of AI requires a platform to meet their needs. Some of these needs include the following:



Market Insight

79% of organizations said they must use AI in business- and mission-critical processes to compete better.³

- **Data abundance and complexity that require scale:** Financial services generate massive amounts of structured and unstructured data daily, such as transactions, market data, customer interactions, and regulatory filings. AI excels at processing and analyzing data in real time to uncover patterns, trends, and actionable insights.
- **Enhanced decision-making:** AI-powered algorithms can forecast market trends, assess credit risks, and predict customer behavior with high accuracy. AI can also automate decision-making processes, such as approving loans or flagging fraudulent transactions, leading to faster, more consistent outcomes.

³ Source: Enterprise Strategy Group Complete Survey Results, [The State of DataOps: Unleashing the Power of Data](#), December 2023.

- **Risk management and fraud detection:** Detecting patterns of financial fraud on a large scale presents challenges to the volume of transaction data that must be processed quickly. Additionally, the limited availability of labeled data for fraud instances obstructs the training of effective models. Banking and payment companies face obstacles in fraud detection, including slow processing times, minimizing false positives, ensuring data integration and quality, and meeting low-latency requirements for real-time decision-making.

Low latency enables the quick assessment of a customer's payment history, enabling the enterprise to detect potential fraud before approving the transaction and reducing the risk of fraudulent activity.

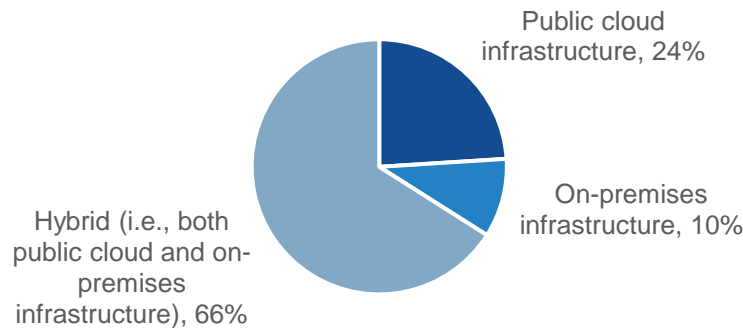
- **Customer experience enhancement:** AI-driven recommendations improve customer experience by tailoring financial products, investment options, and communication to individual needs. Chatbots and AI assistants power customer service tools to provide 24/7 support, answer queries, and automate routine requests. For sentiment analysis, AI proactively analyzes customer feedback to improve products and services.
- **Operational efficiency:** Process automation and AI streamline back-office tasks, such as compliance checks, reconciliation, and report generation. Automating routine processes can reduce costs while maintaining accuracy and scalability.
- **Driving innovation:** AI empowers financial institutions to rapidly develop and deploy new products, creating opportunities for growth and differentiation.
- **Regulatory compliance:** AI helps institutions comply with complex and evolving regulations by monitoring transactions, flagging anomalies, and generating audit trails.
- **Creating competitive advantage:** Financial institutions leveraging AI can innovate faster, make better decisions, and deliver superior services, giving them an edge in a highly competitive and rapidly evolving industry.

Cloudera addresses all of the requirements with its long history of supporting financial organizations with the data platform they need to help them operate at scale and address current and future data-driven use cases, both on-premises and in the cloud.

Hybrid Is a Top Requirement

In the financial world, many organizations have high volumes of data on premises and leverage the cloud for specific workloads. The data platform to support this type of environment requires unique expertise. According to Enterprise Strategy research, organizations were asked about their infrastructure needs for a modern data platform. 66% of organizations reported the need for a hybrid environment with the flexibility to manage their environment on premises and in a single or multiple cloud environment (see Figure 3).⁴

⁴ Source: Enterprise Strategy Group Research Report, [Data Platforms: The Path to Achieving Data-driven Empowerment](#), June 2023.

Figure 3. A Hybrid Infrastructure Is Preferred for Most Organizations**What is your organization's infrastructure preference to support its modern data platforms? (Percent of respondents, N=354)**

Source: Enterprise Strategy Group, a division of Informa TechTarget

Cloudera True Hybrid

Cloudera True Hybrid provides a unified, consistent, and flexible data platform that seamlessly operates across on-premises, public cloud, and private cloud environments. This capability empowers organizations to build and manage advanced data architectures, including enabling data, AI, and analytics across hybrid environments, which is a unique and powerful capability.

Cloudera: Data Platform for the Financial Industry

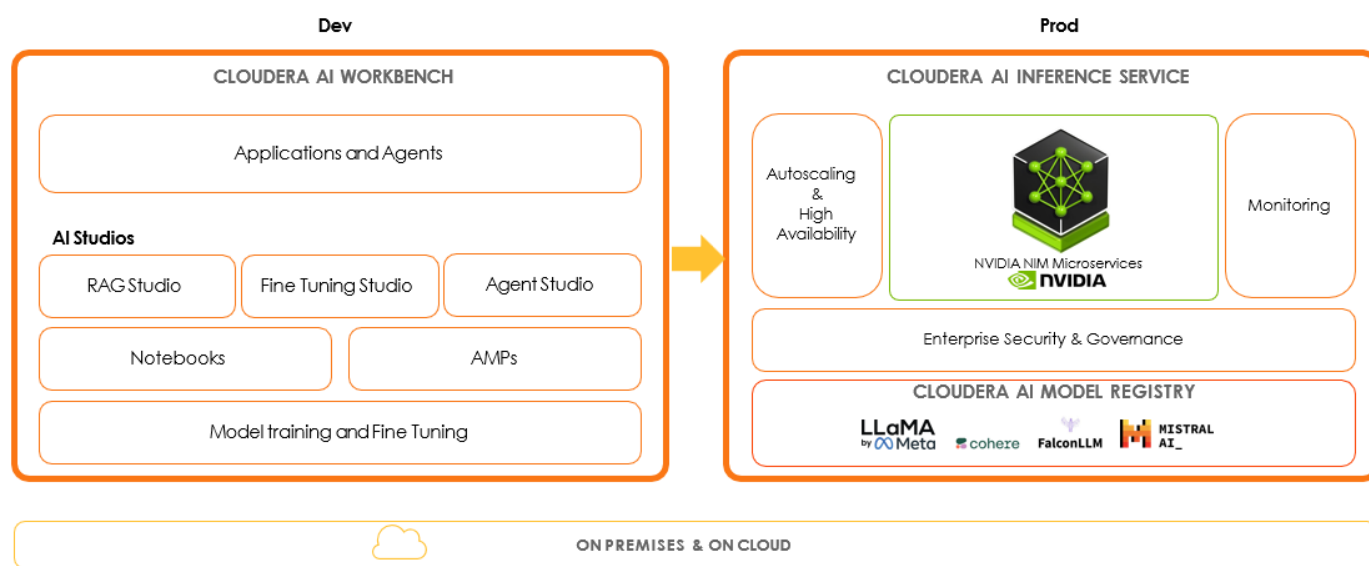
Cloudera centers around building trusted data for organizations that can be democratized across the business with the insights to power decision-making and enable business operations. Some of the main features of the Cloudera platform include:

- **Open data lakehouse:** Cloudera's open data lakehouse design bridges the gap between data lakes and data warehouses and provides scalable and secure data storage for both structured and unstructured data. It enables businesses to easily collect and visualize data in hybrid environments without compromising scalability and vendor lock-in.
- **Self-service analytics:** Cloudera offers self-service analytics, enabling business users and data scientists to view, visualize, and interpret data in their own ways. This data democratization drives faster insights and gives teams the power to make informed decisions without heavily investing in IT.
- **Data transformation and pipelines:** In addition to powerful data transformation and pipeline tools, Cloudera automates and simplifies data preparation for organizations' analysis. From ingest to analysis, the platform facilitates many different data flow scenarios, delivering high-quality data fit for big data and AI use cases.
- **Scale and performance:** Cloudera aims to scale naturally, meeting the needs of even the largest businesses with huge amounts of data. Its hyper-performance design makes it possible for organizations to analyze and consume data in real time to aid mission-critical processes and rapid decision-making.
- **AWS relationship:** In collaboration with Amazon Web Services (AWS), Cloudera leverages AWS to offer its customers cutting-edge analytics and AI tools. This integration offers high performance, savings, and the flexibility to use the entire suite of AWS services as well as Cloudera's hybrid data infrastructure. NVIDIA GPU-based instances are an important part of the AWS EC2 arsenal for accelerated computing featured in Amazon Elastic Compute Cloud (Amazon EC2). They provide a seamless, cost-effective, and highly performant option for deploying Cloudera.

Cloudera AI With NVIDIA for Fast, Effective AI Development

With financial organizations embracing AI, it's critical to have a partner that can deliver outstanding data management capabilities to build trusted data to operate the business. It's also critical to have the ability to use data for AI and analytics with the scale, agility, and performance needed. Cloudera AI provides a fast route to achieving trusted and secure generative and predictive AI and facilitates effortless integration with any language or foundation model for efficient AI development. When an organization has trusted data running on a high-performance platform, it can accelerate its path to using enterprise data to create unique AI-powered experiences. Figure 4 shows some of Cloudera's core development and production capabilities that enable organizations to build and run world-class AI applications and agents.

Figure 4. Cloudera AI Unified Platform



Source: Cloudera

The Cloudera AI workbench empowers developers to utilize their data stored on Cloudera to create test and development environments for AI-powered applications and agents in both on-premises and cloud environments.

Development Environment

These capabilities collectively empower organizations to efficiently develop, deploy, and manage machine learning models within a unified platform that includes:

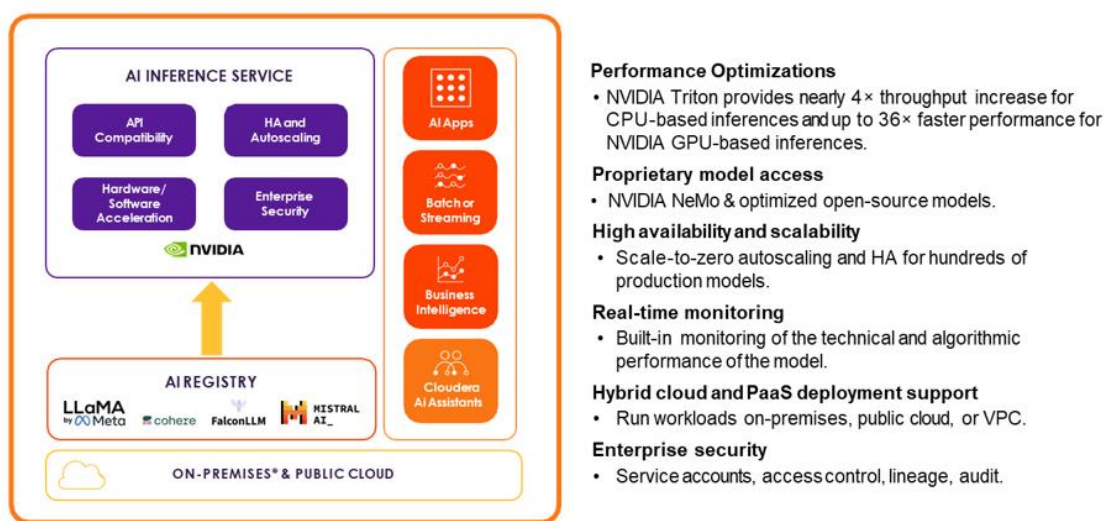
- **RAG Studio:** This function facilitates the creation of retrieval-augmented generation (RAG) applications by integrating knowledge graphs. It enhances the performance of RAG systems by capturing relationships and contexts that are not easily accessible by vector stores alone.
- **Fine Tuning Studio:** An all-encompassing application for managing, fine-tuning, and evaluating large language models, this feature enables users to organize data, design training prompts, train adapters, and assess model performance within Cloudera's AI ecosystem.
- **Agent Studio:** This tool assists in building and deploying AI agents and agentic applications tailored to specific tasks, streamlining the process of creating intelligent agents that can interact with users or systems effectively.

- **Notebooks:** Cloudera provides interactive notebooks that serve as a collaborative environment for data scientists and developers to write, execute, and share code, facilitating exploratory data analysis and model development.
- **Accelerators for Machine Learning Projects (AMPs):** These are prebuilt, end-to-end machine learning projects that can be deployed with a single click. Developed by Cloudera's Fast Forward Labs, AMPs enable data scientists to rapidly prototype and deploy ML use cases, reducing time to value.
- **Model Training and Fine Tuning:** Cloudera's platform supports the training and fine-tuning of machine learning models, enabling users to adapt pretrained models to specific tasks or data sets, thereby enhancing model performance and relevance to particular use cases.

Production Environment

The product environment is shown in Figure 4, with an expanded view of Cloudera AI Inference in Figure 5. Inference is the process of using the trained models in production to make predictions or decisions based on new, unseen data. Here, the model applied the knowledge it learned during training to perform tasks such as classification, regression, object detection, and natural language processing.

Figure 5. Cloudera AI Inference



Source: Cloudera

The Cloudera AI production environment provides the following critical deployment capabilities:

- **NVIDIA NIM Microservices:** NVIDIA NIM plays an important role in optimizing inference performance, which is a critical function for AI success. These microservices leverage the NVIDIA AI platform to accelerate parallel processing, enabling efficient handling of large data volumes and complex computations. Designed to integrate seamlessly with Cloudera's AI platform, they support various deep learning and machine learning frameworks, ensuring that developed models can be deployed and scaled efficiently in production environments.
- **Auto scaling and high availability:** These fundamental features ensure resilience and efficiency. Auto-scaling dynamically adjusts computational resources based on workload demands, scaling up during peak periods and scaling down during idle times to optimize costs. High availability ensures continuous operation by distributing workloads across redundant systems, minimizing downtime, and maintaining service continuity even during system failures through robust failover mechanisms.

- **Monitoring:** With all applications, monitoring plays an important role in providing real-time visibility into system and model performance, tracking key metrics such as latency, throughput, resource utilization, and error rates. This capability helps organizations detect issues proactively, enabling timely alerts and troubleshooting to maintain consistent production AI workflow and adherence to service-level agreements.
- **Enterprise security and governance:** Safeguarding sensitive data and AI models in production is a core requirement for AI development. The platform enforces robust security protocols, including role-based access control (RBAC), to ensure that only authorized personnel can manage or access AI systems. It supports enterprise governance policies with audit trails, data usage monitoring, and encryption for data both in transit and at rest, protecting against unauthorized access or breaches. These features ensure compliance with regulatory and organizational security standards.
- **The Cloudera AI Model Registry:** This centralized repository is designed to give users access to AI models throughout their lifecycle. It provides version control to ensure traceability and rollback options, enabling seamless deployment from the registry to the inference environment.

By optimizing for inference, financial organizations will be able to quickly respond to market fluctuations with higher throughput and performance and deliver improved customer service with sophisticated generative AI agents. Cloudera's unified platform enables financial organizations to have an all-in-one solution for creating, deploying, and managing custom AI solutions using enterprise data with the integrated performance of Inference Service with NVIDIA NIM microservices, which ensures autoscaling, high availability, and enterprise-class security and governance in production.

Private AI for Financial Firms

Among financial firms, there is a growing demand for Private AI. Cloudera and NVIDIA are uniquely qualified to address this demand given their ability to address AI workloads running on premises and in the cloud. Private AI gives organizations a high level of security and control over data, large and small language models, and the deployment of new AI-powered applications and models managed across public clouds and on premises.

Simplified Deployment and Support

Another advantage of Cloudera AI Inference with NVIDIA is centered around simplicity, with a one-click deployment, one integrated security solution, one single platform to manage with Cloudera and integrated NVIDIA NIM, and one place to get the support needed.

Figure 6. Ease of Getting Started With Cloudera



Source: Cloudera

Conclusion

The Cloudera-NVIDIA partnership enables financial organizations to gain access to an innovative, high-performance data platform that can accelerate AI innovation. The Cloudera true hybrid model is also unique in its ability to extend capabilities across on-premises environments to public clouds, providing financial organizations full control over their data, AI models, and applications. Cloudera's cutting-edge data, AI, and analytics capabilities are designed to perfectly suit the needs of financial organizations and provide the scale and confidence needed to facilitate diverse AI use cases, leveraging the power of NVIDIA for performance, scale, and security to power the perfect foundation for success. Such synergy enables financial organizations to harness the full potential of their data, leading to faster innovation and better business results in a highly competitive market.

Enterprise Strategy Group highly recommends Cloudera's all-in-one AI solution for financial services organizations striving to respond to increasing workloads and remain competitive in the data-driven world.

©TechTarget, Inc. or its subsidiaries. All rights reserved. TechTarget, and the TechTarget logo, are trademarks or registered trademarks of TechTarget, Inc. and are registered in jurisdictions worldwide. Other product and service names and logos, including for BrightTALK, Xtelligent, and the Enterprise Strategy Group might be trademarks of TechTarget or its subsidiaries. All other trademarks, logos and brand names are the property of their respective owners.

Information contained in this publication has been obtained by sources TechTarget considers to be reliable but is not warranted by TechTarget. This publication may contain opinions of TechTarget, which are subject to change. This publication may include forecasts, projections, and other predictive statements that represent TechTarget's assumptions and expectations in light of currently available information. These forecasts are based on industry trends and involve variables and uncertainties. Consequently, TechTarget makes no warranty as to the accuracy of specific forecasts, projections or predictive statements contained herein.

Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of TechTarget, is in violation of U.S. copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact Client Relations at cr@esg-global.com.

About Enterprise Strategy Group

TechTarget's Enterprise Strategy Group provides focused and actionable market intelligence, demand-side research, analyst advisory services, GTM strategy guidance, solution validations, and custom content supporting enterprise technology buying and selling.

✉ contact@esg-global.com

🌐 www.esg-global.com