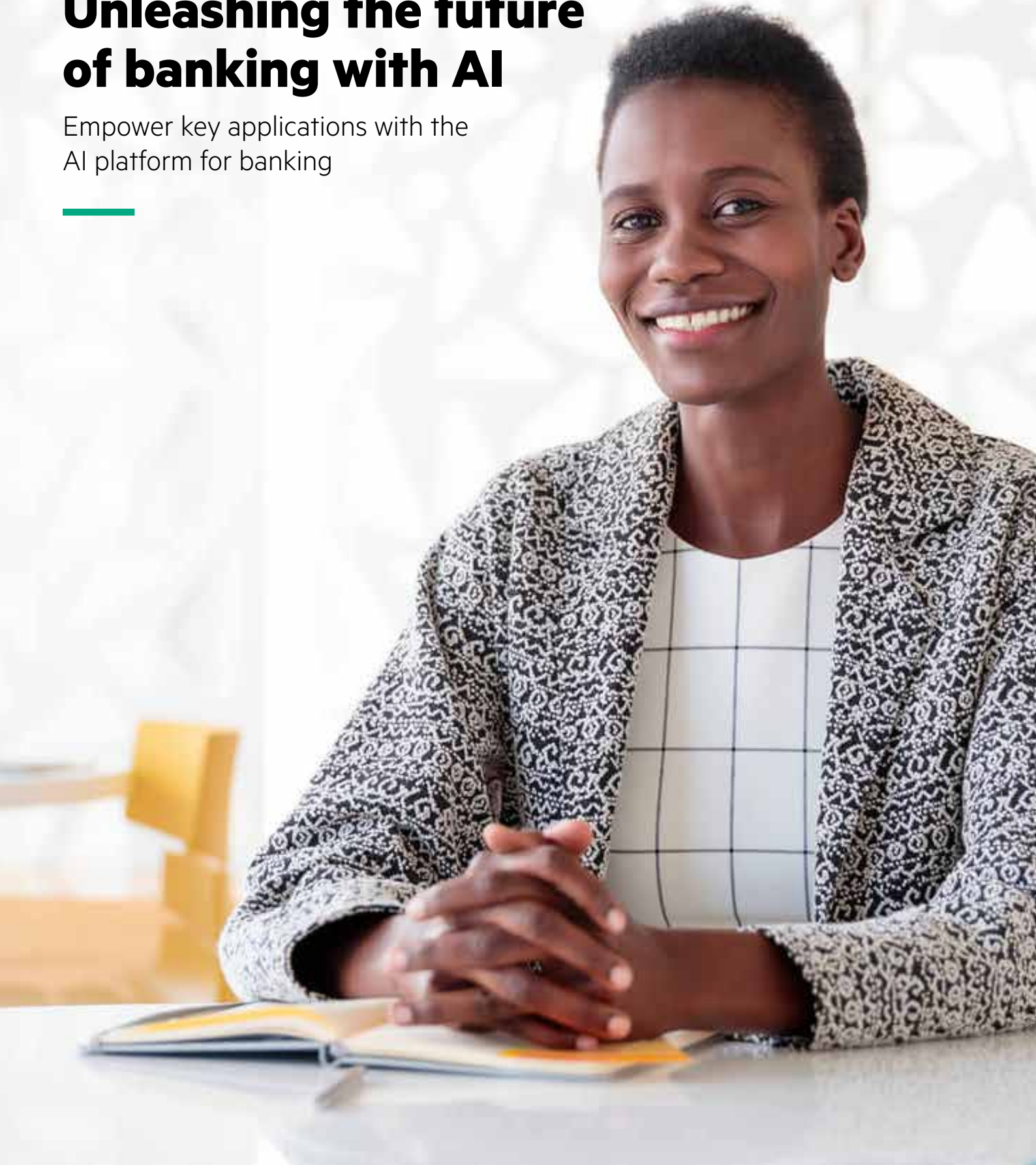# Unleashing the future of banking with AI

Empower key applications with the
AI platform for banking

Operationalize AI to gain faster, broader insights
Banks understand the benefits of adding data analytics, AI, and ML to their digital initiatives. However, as they try to implement these technologies, they quickly discover the large gap that exists between the amount of data they have and how much of it can be analyzed to create business value. HPE and NVIDIA® make it faster to operationalize AI/ML with an AI platform for banking that can be configured and scaled to meet any requirements.

The new age of banking

Data is the most valuable currency in the banking industry. Today's institutions have access to a wealth of data that is being generated from sources like customer information, customer engagements, customer feedback, sales, transactions, spending patterns, mobile applications, and daily activities. However, all this data means nothing without insights: the capacity to gain an accurate and deep understanding of your business and customers. Banks are racing to uncover real-time and actionable insights from data that can help them develop competitive new services, boost customer satisfaction, and drive efficiencies. In fact, banking is one of the top three industries that make the largest annual investments in big data and analytics. Institutions that run on legacy technologies are quickly reaching their limits as data volumes grow exponentially. The inability to distinguish and utilize valuable data leads to missed business opportunities and lost profits. Advancements in artificial intelligence (AI) enable banks to increase the speed, accuracy, and productivity of their operations, providing key capabilities for them to innovate and grow.

These demands have put tremendous pressure on institutions to transform. To prepare for the future, banks are making improvements in six main areas:

• **Resiliency:** Resiliency will be the new modernization goal. Everything from fraud prevention, risk management, cloud, and data will be included in this initiative.

• **Digital payments:** Already a driver, digital payments will grow in importance, especially as more customers demand anytime, anywhere, and contactless services.

• **Security:** As the industry integrates new security architectures to enhance data protection, banks will reconsider authentication and fraud prevention practices across the enterprise.

• **Scoping AI projects:** Banks will need to determine appropriate AI use cases and work to create and refine accurate AI / machine learning (ML) models.

• **AI funding:** AI-enabled applications improve operations, enhance customer satisfaction, and drive top-line revenue. Banks will allocate more of their technology budgets to deploy and manage effective AI tools.

• **IT infrastructure:** Banks must invest in scalable, high-performing AI platforms that are reliable and secure to meet the requirements of AI across the enterprise and efficiently run AI algorithms and models at scale.

To address these new requirements, banks must find ways to rapidly convert data into insights, when and where they are needed most. According to Forrester, 98% of institutions believe MLOps practices—that is, collaboration between operations professionals and data scientists to effectively achieve ML model lifecycle management—will give them a competitive edge. Specifically, they expect an increase in ML capabilities to enable 53% higher profitability as well as 52% better customer experiences. Unfortunately, as many as 85% of ML projects fail to reach production. Banks are running into the last-mile problem with AI deployment and management, which affects 80–85% of enterprises. The reasons AI projects fail have little to do with having the best data science team and more to do with standard business and IT integration challenges.

Integrating an enterprise AI platform for banking will be essential to succeed. The right solutions will put performance and intelligence at the core of banking operations to enable insights on demand and prepare institutions to tackle today's and tomorrow's challenges.

Advance your AI and deep learning expertise

Drive deep learning (DL) and AI adoption at all stages throughout the enterprise with GPU-accelerated compute and consultative expertise in AI solution development from HPE and NVIDIA. Build expertise through online and instructor-led workshops, reference architectures, and benchmarks on NVIDIA GPU-accelerated applications and enhance time to value. Accelerate real-time insights.

Rethink banking with an end-to-end platform for AI operations

Banks need partners they can trust to help them deploy and manage effective AI infrastructure. In response to the mounting challenges faced by banks as they continue along their AI journey, Hewlett Packard Enterprise and NVIDIA have partnered up to offer a platform for enterprise AI operations. Instead of trying to specify compute and storage needs for an unpredictable set of requirements and demands in the future, HPE software and HPE Reference Architecture for AI-Optimized Infrastructure with NVIDIA GPUs and software allow banks to quickly deploy the resources you need now, with the option to easily expand capacity as you need it. This solution provides a self-contained high-performance AI infrastructure including the compute, storage, network, and software elements to handle the most complex AI challenges when developing, deploying, and maintaining multiple AI applications for banking.

In partnership with NVIDIA, HPE makes it seamless for banking line of business executives, data scientists, and IT professionals to work together by addressing their unique workloads and requirements and providing the AI capabilities they need to produce and share insights faster. We know an effective solution requires more than just performant storage and computational power—banks benefit from a robust software layer and infrastructure that is flexible, scalable, and can support applications with mixed requirements. That is why we offer the HPE AI platform for banking featuring NVIDIA technologies and tools, engineered to help you rapidly innovate, operate more effectively, and realize better business outcomes—all while driving down your TCO.

Our AI platform for banking features accelerated application frameworks, pre-trained models, tools, libraries, and SDKs for application developers to quickly build AI algorithms for competitive new services. Banks can also choose from numerous open source, NVIDIA-Certified AI applications from a diverse ISV ecosystem to push their productivity to new heights. This high-performance compute infrastructure can be applied across multiple dimensions of AI complexity—data volumes, AI/models, team size, and environment from edge to cloud.

AI-powered banking applications

The future of banking will be data-driven, delivered digitally, and secured in a hybrid cloud environment. Banks that make AI a priority will be able to jumpstart their transformation across all areas of operations.

AI provides the extreme agility and precision that are required to achieve faster insights. These capabilities make it easier to utilize vast quantities of data quickly and securely. Banks can rapidly monetize their data by taking steps to mitigate risk, ramping up services, and improving the customer experience. To make this possible, AI technologies are driving progress in four main banking applications—fraud detection and identity verification, conversational AI, robotic processing automation (RPA) for document processing, and recommendation engines.

These workloads utilize AI in unique ways and therefore have different data processing and storage requirements. To run each application effectively, banks need a solution that offers simplicity, flexibility, and scalability to accommodate heterogenous workloads and architectures. HPE and NVIDIA deliver a platform that can run each AI workload on the same integrated infrastructure, including the most common types of workloads.

Fraud detection and identify verification
Reducing risk is top of mind for banking executives. Security, regulatory compliance, anti-fraud, anti-money laundering (AML), and know your customer (KYC) guidelines are included under the umbrella of risk management. In 2020, banks worldwide amassed $15.13 billion in fines due to cyberattacks and fraud. U.S. banks accounted for more than 73% of the issued fines, totaling $11.11 billion. Fraud detection and identity verification

are critical to enhance data security and reduce risk. Powered by AI, banks can use these applications to calculate risk and fraudulent exposure in real time. Banks are employing AI to seamlessly authenticate users and identify unusual behavior, allowing them to scan payments, isolate transaction patterns, and flag anomalies in milliseconds.

Conversational AI

Conversational AI and automatic speech recognition are streamlining processes in the call center. Instead of spending several minutes documenting customer engagements and feeding information back into the system following events, AI can deliver a better customer support experience by using natural language inputs. Together, conversational AI and automatic speech recognition applications analyze video and audio content instantly, streaming insights from every customer interaction. This helps to accelerate and automate daily processes and enables banks to promptly respond to customer needs.

Applica, a leading ISV of robotic text automation solutions, offers NVIDIA-Certified applications that combine computer vision and DL-driven natural language processing for all document types. The software uses NVIDIA GPUs to train AI models and inference in production, enabling banks to interpret information and make informed decisions in seconds while eliminating up to 90% of errors. Applica's solutions are transforming key banking use cases including mortgage origination and KYC. With the ability to extract any data point, regardless of the document format or complexity, mortgage lenders can close loans 12 days quicker and increase their Net Promoter Score by 7%. The increased automation accuracy also enables organizations to avoid manual KYC processes that are time-consuming and costly, decreasing the turnover time from two days to just five minutes and elevating customer experiences.

Robotic process automation

RPA is used to extract meaning from a wide range of documents, regardless of the language or layout, and process workloads with unstructured document types. This application does much more than simply process documents. Banks employ sophisticated AI analysis tools to automate document processing workflows, making it faster to interpret documents and make human-like decisions. This eliminates much of the labor and costs of processing events manually.

Recommendation engines

Banks employ recommendation engines to determine the next best action. Using AI and ML models, recommenders offer personalized suggestions based on information provided by customers. The models analyze large amounts of data—such as credit scores, outstanding balances, account utilization, and other personal details—to detect early warning signs of risk including defaults, bankruptcies, fraud, and lawsuits. From the collected insights, recommendation engines enable banks to efficiently dispatch resources and increase customer conversion rates. The actions help to improve loyalty and provide a seamless banking experience.

Capitalize on AI applications

To help engineers, programmers, and data scientists take advantage of these ubiquitous AI technologies as well as those that are new and innovative to the market, HPE and NVIDIA provide the HPE Ezmeral Marketplace and the NVIDIA GPU Cloud (NGC) respectively to access third-party software that can augment your suite of AI tools and applications.

The HPE Ezmeral Marketplace leverages the open source Kubernetes ecosystem and its modern ISV partners to provide faster time to value. Together, they make it easier for organizations to adopt integrated solutions that combine HPE Ezmeral software with validated industry-leading, third-party, commercial, and open source applications.

With the HPE Ezmeral Marketplace, ISVs validate their software on the HPE Ezmeral software, constantly adding functionalities to the marketplace through other vendors that include big data and analytics, AI/ML, data protection, monitoring, security, and anything else needed for the end-to-end data pipeline. Through the HPE Ezmeral Marketplace, banks can fill in the gaps without getting boxed into just one set of offerings, recognizing that no enterprise is using just one tool to meet all their needs.

The NGC Catalog is a curated set of GPU-optimized software for AI, HPC, and visualization. The content provided by NVIDIA and third-party ISVs simplify building, customizing, and the integration of GPU-optimized software into workflows, accelerating the time to solutions for users. The NGC Catalog consists of containers, pre-trained models, Helm charts for Kubernetes deployments, and industry-specific AI toolkits with software development kits (SDKs). Major SDKs will be available soon via NVIDIA AI Enterprise, making AI accessible to organizations of any size. NVIDIA AI Enterprise is an end-to-end, cloud-native suite of AI and data analytics software that is certified to deploy anywhere and includes global enterprise support to keep AI projects on track.

The HPE AI platform for banking

Despite the operational and competitive advantages of AI, many banks lack the tools or expertise to unlock its full potential. Banks must develop an effective AI strategy if they want to redefine what's possible for banking. Those that do will develop a robust technology infrastructure that can support AI-powered applications, allowing them to gain broader insights and scale for the future.

For banks that have committed to leveraging AI and are running a variety of AI POCs or have models in production across the enterprise, the HPE AI platform for banking provides an open, cloud-like platform that is hybrid by design to help you better manage your AI/ML architecture and software suite across your institution. Here are the elements of the underlying platform to support your home-grown, open source and/or third-party AI applications now and in the future:

- **HPE Ezmeral software suite**

  – HPE Ezmeral MLOps

  – HPE Ezmeral Runtime Enterprise

  – HPE Ezmeral Data Fabric

  – HPE Ezmeral Unified Analytics

- **HPE Reference Architecture for AI-Optimized Infrastructure**

  – **Compute:** HPE Apollo 6500 Gen10 Plus with NVIDIA A100 GPUs

  – **Networking:** Compute/storage fabric, in-band management network with NVIDIA Quantum InfiniBand, and out-of-band management network

  – **Storage:** Cray ClusterStor, Parallel File System, HPE Ezmeral Data Fabric File and Object Storage

- **Professional services and expertise**

  – HPE Pointnext Services

  – HPE GreenLake Cloud Services

HPE Ezmeral software suite

The HPE Ezmeral software suite creates a modern infrastructure with frictionless access to hybrid data. Your teams can easily infuse analytics, AI, and ML into every data workflow, IT process, and business decision. HPE Ezmeral MLOps enables effective utilization of resources, so data scientists can focus on building models and analyzing results, rather than waiting for training jobs to complete. To provide a consistent experience across teams, HPE Ezmeral Runtime delivers one platform with a wide variety of analytics, AI, and ML tools—everything your key end users need to be productive.

To overcome data silos, disconnected data teams, and a lack of standards that are making it difficult to achieve the next level of innovation, HPE Ezmeral Unified Analytics provides a unified data experience that allows teams to securely connect to data where it exists today, without disrupting existing data access patterns.

Institutions are already seeing significant competitive advantage with HPE Ezmeral software:

• More than $7 million in infrastructure savings over three years

• More than 1200x faster time to value

• 33% higher productivity for analytics teams

• 567% ROI over five years

From this comprehensive software layer, the HPE AI platform for banking combines industry-leading hardware and support services to create one tightly integrated solution that can help banks manage the requirements of AI as well as keep pace with evolving industry trends and regulations.

HPE Reference Architecture for AI-Optimized Infrastructure
Aligned with the NVIDIA reference architecture, the HPE Reference Architecture for AI-Optimized Infrastructure provides a high-performance compute, storage, network, and software elements to handle the most complex AI business problems, working with NVIDIA NGC-optimized software.

AI resource demands will only ever increase. With this AI-optimized infrastructure, the expansion process is quick and simple. Additional scalable units (SUs) within the integrated system can be deployed in a matter of hours, delivering a large resource boost without affecting existing AI operations. In addition to NVIDIA A100 GPUs, this platform delivers superior performance using proven technologies including HPE Apollo 6500 Gen10 Plus nodes, 200 Gb/s InfiniBand connectivity, ClusterStor E1000 storage, HPE Parallel File System Storage, and HPE Ezmeral Data Fabric File and Object Store.

HPE Apollo 6500 Gen10 Plus systems are purpose-built as the HPE Reference Architecture for AI-Optimized Infrastructure and validated and powered by NVIDIA A100 HGX™ GPUs. The infrastructure is inherently scalable and designed to run AI between nodes as well as within nodes. NVIDIA-Certified systems integrated with NVIDIA ConnectX-6 adapters provide performance acceleration such as network and storage offload, security, and remote management capabilities. ConnectX-6 adapters ensure the best node-to-node connectivity and scalability of your AI workload.

AI-optimized storage solutions from HPE are capable of managing your teams' storage AI and analytics demands. HPE Ezmeral Data Fabric File and Object Store simplifies data management by unifying any data type from a wide variety of technologies into a single data infrastructure and file system that scales as needed without impacting service levels, security, or data resilience. HPE Ezmeral Data Fabric is the industry's first edge-to-cloud data fabric purpose-built for analytics and offers key advantages over other object-based solutions. This modern data fabric gives users the ability to run multiple AI use cases on a single cluster and provides maximum storage for various dataset structures. This helps to streamline the deployment of different AI-driven applications and data pipelines into production using one architecture. Banks benefit from extreme agility and increased reliability by optimizing all object sizes for both performance and storage efficiency in a persistent datastore to accelerate analytics.

Together, HPE and NVIDIA deliver accelerated compute that is capable of powering AI development, pre-production, and production in the same physical infrastructure. The robust architecture enables banks to operate at unprecedented speed and scale, running a single massive AI workload or eliminating inefficiencies across disparate locations. Banks can scale their AI platform as data volumes and processing requirements increase, offering a high degree of future-proofing with the superior performance, resilience, and security of trusted HPE and NVIDIA hardware and proven security features. We build security into the core of every AI platform. From the start of production, the HPE Trusted Supply Chain provides a new first line of defense against cyberattacks and delivers the end-to-end security that banks demand. Hardened data protection during the manufacturing process helps to defend sensitive applications and customer details before HPE systems are even built.

The HPE silicon root of trust creates an immutable fingerprint in the silicon that prevents systems from booting with compromised firmware and automatically restores the previous uncompromised firmware, which makes HPE systems the world's most secure industry-standard servers. These breakthrough technologies drive enterprise security from production and distribution to deployment, whether banks choose to utilize our solutions on-premises, in the cloud, or in a containerized environment.

Professional services and expertise

HPE Pointnext Services and HPE GreenLake Cloud Services help simplify AI and boost operational flexibility, so organizations can innovate faster. The vast array of professional support and services enable banks to easier their adoption of HPE AI Optimized Cluster solutions. HPE Pointnext Services help banks design and implement a successful AI strategy, from pilot to production. Combined with the AMP assessment program, our experts work with you to speed up Kubernetes adoption and kickstart your application modernization process.

Whether a complete AI platform or cloud services for MLOps, HPE GreenLake can help you get started and even manage the solution for you. HPE GreenLake gives banks choice in how to deploy and manage AI. Cloud services for MLOps make it faster to adopt AI/ML and seamlessly scale into production. With enterprise-grade container management services and pre-integrated data fabric, you can run Kubernetes at scale on-premises to streamline app-driven innovation.

Conclusion

HPE and NVIDIA are delivering comprehensive AI solutions that promise to empower banks throughout their AI journeys. The HPE AI platform for banking with NVIDIA technologies is the next stage of innovation that will prepare institutions for greater insight and productivity. We are dedicated to helping you succeed and equipping each of your teams to work with greater performance and intelligence.

Let us help you unleash the potential of AI for banking. Contact HPE today to begin your transformation.

Make the right purchase decision. Contact our presales specialists.

Chat now (sales)

Call now

Learn more ▭

hpe.com/partners/nvidia
hpe.com/info/ai

Get updates

Hewlett Packard Enterprise