

AI Empowers Innovative Banks

Drive better banking outcomes with AI
solutions from Dell Technologies and NVIDIA



Table of Contents

The age of AI is here 3

Primary use cases for the AI-empowered bank 4

 Ramping up AML and identity verification 5

 Improving transactional fraud prevention 6

 Enhancing CX cost effectively with digital assistants and chatbots 8

 Activate AI-empowered banking outcomes for all your use cases 8

Simplify and speed AI with Dell Validated Designs 9

 Know what’s under the hood 10

Establish a flexible foundation with Dell APEX-as-a-Service solutions 12

 Modernize for AI by expanding your use of cloud 12

 Dell APEX 13

Partner with Dell Technologies and NVIDIA experts 14

 HPC & AI Innovation Lab 14

 HPC & AI Centers of Excellence 14

Services and financing for AI 15

Activate better banking outcomes with AI 15

 Learn More. 15

Appendix 16

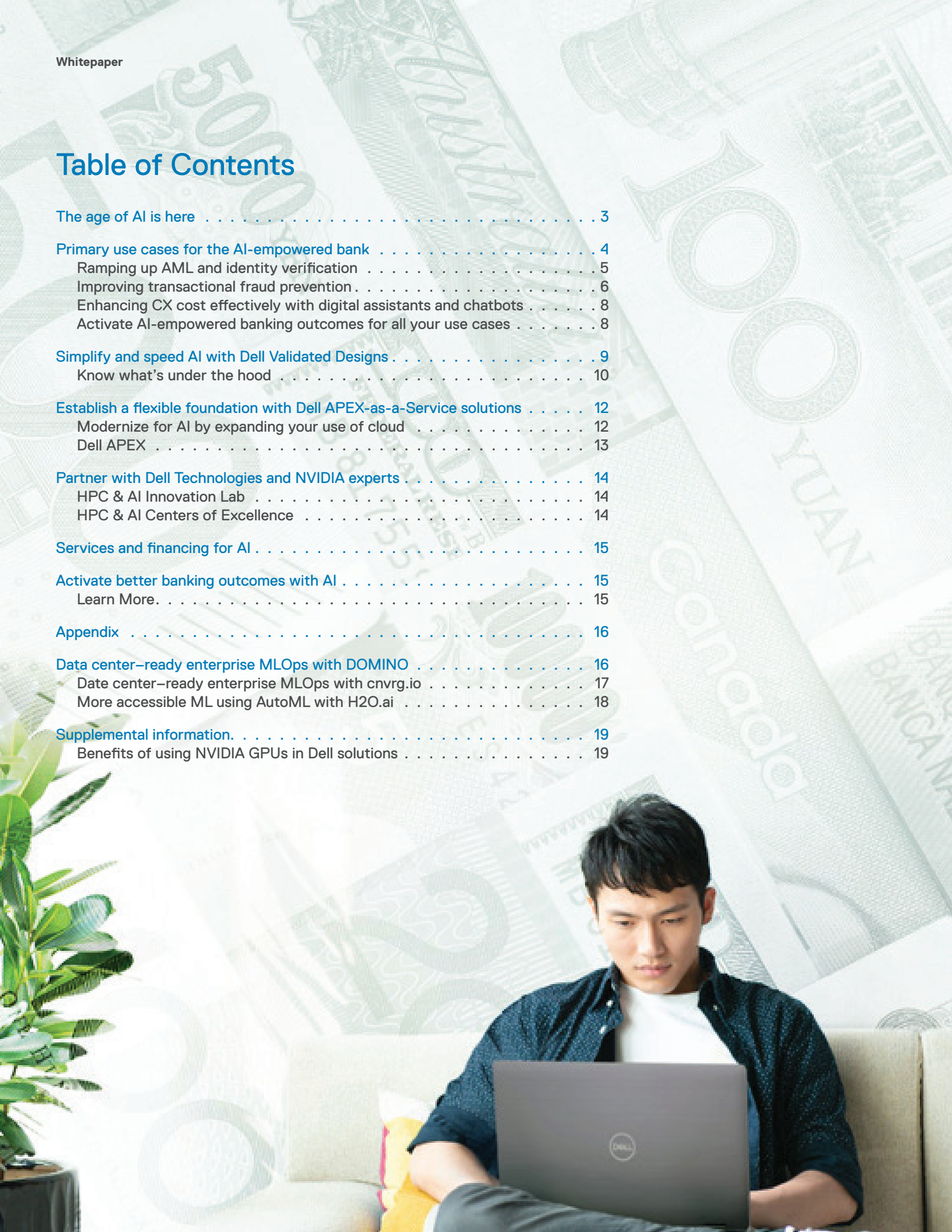
Data center–ready enterprise MLOps with DOMINO 16

 Date center–ready enterprise MLOps with cnvrg.io 17

 More accessible ML using AutoML with H2O.ai 18

Supplemental information. 19

 Benefits of using NVIDIA GPUs in Dell solutions 19



The age of AI is here

Artificial intelligence (AI) can discover patterns in data that are simply too complex for humans — even experts — to recognize. AI is also adaptable, scalable and essentially tireless. However, not very long ago, developing and implementing an AI solution meant making an enormous investment in platforms, tools, people and time. Such an endeavor was only accessible to the largest of financial institutions. Today, thanks to advances in technology, AI is readily available and accessible for financial institutions of all sizes, using solutions that can be built and deployed quickly, securely and cost effectively.



Welcome to the age of AI

94%

of business leaders agree that AI is critical to success over the next five years.¹

76%

of business leaders report they plan to increase investments in AI to gain more benefits.¹

50%

of business leaders face significant challenges to scaling AI across the business.¹

¹ Deloitte, [Fueling the AI transformation: Four key actions powering widespread value from AI, right now](#), October 2022.

AI transforms the way banks operate. By helping identify key insights in vast amounts of data, calculate risk and automate routine tasks — all at unprecedented speed and scale — every line of business and function can be enhanced. This enables the AI-empowered bank to increase productivity, expand services, reduce risk and dramatically improve customer service. The result is increased revenues, decreased costs and the ability to stay a step ahead of the competition by executing on real-time insights garnered from enormous amounts of data.

Outcomes for the AI-empowered bank

- Increased revenues
- Decreased costs
- Real-time insights

Primary use cases for the AI-empowered bank

Ultimately, AI-empowered banks will lead the market for financial services into the future, as they deliver enhanced AI applications to internal stakeholders and external customers that create competitive advantage in the market. Leading banks are already operating against hundreds of AI projects. In the near future, thousands of applications across the bank will be AI-enabled.

Some of the more prevalent use cases for AI-empowered banks include anti-money laundering (AML) and identity verification, transactional fraud prevention, digital assistants, natural language processing (NLP), transformer models, computer vision, high performance computing (HPC), analytics, recommendation engines and digital twins. These workloads support a variety of mission-critical needs including customer service; fraud detection; environmental, social and governance (ESG); risk management and more.

In this paper, we will focus on the top banking use cases:



AML and identity verification



Transactional fraud prevention



Digital assistants / chatbots



“For global banking, McKinsey estimates that AI technologies could potentially deliver up to \$1 trillion of additional value each year.”²



² Quantum Black AI by McKinsey, [The Executive's AI Playbook](#), accessed December 2022.



The United Nations estimates that up to \$2 trillion USD is laundered around the world each year.²



Ramping up AML and identity verification

Money laundering is a serious crime that can have huge consequences for businesses and individuals around the world. According to estimates by security agencies and the financial industry, between US\$800 billion to US\$2 trillion is laundered around the world each year — equivalent to as much as 5% of global gross domestic product (GDP).³

As the volume of money laundering and other financial crimes continues to grow around the world, so does the sophistication of the techniques used by criminals. According to McKinsey, “the result has been a vigorous response from banks, which are collectively investing billions of dollars each year to improve their defenses against financial crime.”² Global spend on AML and know your customer (KYC) data and services is projected to total a record \$1.56 billion in 2022 with a five-year industry compound annual growth rate (CAGR) of 15%.⁴

Many organizations are turning to AI to accelerate analytics and boost accuracy for AML and KYC initiatives as part of identity verification protocols. Previous approaches used rule- and scenario-based tools or basic statistical approaches for transaction monitoring, driven primarily by industry red flags, basic statistical indicators and expert judgment. However, rules often fail to capture the latest trends in money laundering behavior, keeping banks perpetually a step behind criminals and making the fight against money laundering an ongoing challenge for compliance, monitoring and risk organizations.

In contrast, AI algorithms driven by advanced machine learning (ML) models and running on powerful HPC systems can build sophisticated algorithms based on granular data analytics. They are also more flexible, able to quickly adjust to new trends and continually improve over time. Adding deep learning (DL) techniques such as graph neural networks, computer vision and NLP can reduce false positives in transactional fraud detection and enhance identity verification in compliance with AML and KYC requirements, improving both the customer experience and the bank’s financial health. Meaningful gains can be achieved through AI as even minor improvements in detection accuracy significantly reduce costs and improve regulatory compliance. This is leading executives across the industry to integrate AI, analytics and HPC into their current processes.

AI, ML and DL driven analytics along with HPC enable you to deploy the transaction monitoring element of these programs more quickly and easily. With an AI-empowered bank solution, you now have a chance to change the game and stay a step ahead of crime. Dell Technologies can help; with Dell Validated Designs for AI powered by NVIDIA® GPUs and NVIDIA AI Enterprise Suite, there is an opportunity for you to get out in front of AML faster, easier and more cost effectively.

³ United Nations Office on Drugs and Crime, [Money Laundering](#), accessed December 2022.

⁴ Yahoo! Finance, [Global AML/KYC Spending Projected to Total \\$1.6 Billion In 2022, Rising 15% as Governments Continue to Focus on Fighting Financial Crime - New Burton-Taylor Report](#), October 2022.

Improving transactional fraud prevention

Digital financial services interactions are growing exponentially — from mobile banking to e-commerce transactions to online applications for financial products. Traditional banking processes are stretched just trying to sustain this proliferation of digital services, let alone fight the new types of fraudulent activities being perpetrated by bad actors. With online fraud losses expected to reach \$48 billion annually by 2023⁵, improved fraud detection and prevention is the top use case for AI in the financial services industry.

As highlighted in the context of AML and identity verification, detecting fraud can be complicated. The same is true for transaction fraud, as criminals always seem to have a new scheme. Models based strictly on rules become obsolete and could cost financial institutions valuable time and money to update. Further, it can be difficult to strike the right balance between missing some fraudulent transactions due to false negatives and inconveniencing customers with false positives. At the same time, customers expect transactions to go through in an instant. Acting in milliseconds is critical to catching fraud in the act, not after the fact.

AI and ML are capable of performing analytics on massive amounts of data in milliseconds to understand and apply rules. These powerful capabilities require an accelerated computing platform to train increasingly accurate fraud detection AI models and to run more sophisticated ML and DL models in production near instantaneously.

However, when everything is running at speed, AI can make a significant impact on your bottom line. Let's look at an example from Mastercard®.

Customer success: Mastercard

While fraud is a challenge for any company involved in retail sales, the scale at which Mastercard operates makes the problem all the more challenging. Mastercard has 2 billion cards in use, distributed across more than 210 countries and territories. They need to:

- Process 165 million transactions per hour.
- Apply 1.9 million rules per transaction to prevent fraud.
- Do all this in milliseconds to prevent fraud in real time.



Innovating with data to fight fraud in real time

2 billion

cards in 210 countries and territories years.

165 million

transactions per hour

1.9 million

rules examined per transaction

Below 30ms decision-making

rules examined per transaction

Powered by Validated Designs
Dell Data Storage

“Our goal is to stop fraud in its tracks without disrupting or delaying legitimate transactions. . . We want to move intelligence closer to the edge to get closer to our customers.”

Nick Curcuru, Vice President —
Big Data Practice, Mastercard





The name of the game for Mastercard is to outsmart some really smart people who have criminal intent. However, Mastercard needs to stop fraud in its tracks without causing delays in processing credit card transactions, which would negatively affect the customer experience. So rather than processing everything centrally, Mastercard moved intelligence closer to the edge to help their customers get what they're looking for faster and easier.

Mastercard innovates with data in different ways to fight back against fraud every second of every day. To identify and stop fraudulent transactions, Mastercard leverages AI and ML running on HPC systems to analyze large data sets at lightning-fast speeds on a secure, payment card industry (PCI)-certified system from Dell Technologies. This fraud detection ML system uses supervised learning to look for established fraud patterns and unsupervised learning to identify emerging fraud patterns in real time. With every transaction, the algorithms examine things like a cardholder's buying habits, geographic location and travel patterns, along with real-time data on card usage — such as what they are trying to buy, where they're trying to buy it and what else they bought in the same day. Each transaction is analyzed in terms of the rules that relate to what a valid transaction looks like and what a fraudulent transaction looks like.

The end result of these efforts is a more trustworthy transaction experience for legitimate cardholders and merchants and real-time barriers to stop criminals who try to exploit vulnerabilities in the payment systems.

Fighting fraud at Mastercard

ML algorithms running on Dell Validated Designs allow Mastercard to apply 1.9 million fraud detection rules — nearly instantaneously — to 165 million transactions per hour.⁶



⁶ Dell Technologies white paper, [Fighting fraud the smart way — with data analytics and artificial intelligence](#), September 2020.

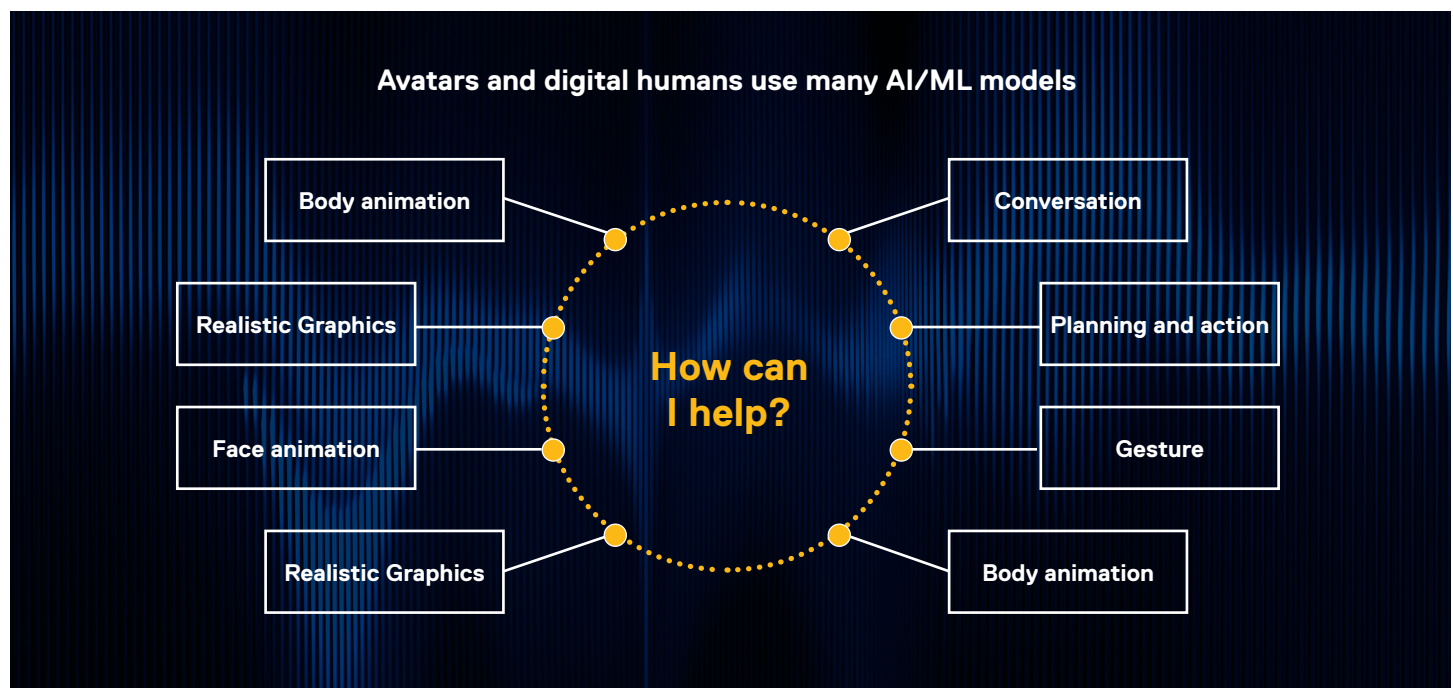
Enhancing CX cost effectively with digital assistants and chatbots

Businesses that prioritize customer experience (CX) grow revenues 1.7x faster and multiply lifetime customer value by 230%⁷, yet optimizing CX has never been more complicated. The opportunity for NLP and conversational AI to impact banks extends beyond the call center into automated customer service channels such as virtual assistants and chatbots. The integration of AI chatbots in banking is on track to save banks 826 million hours and \$7.3 billion in 2023.⁸ AI-driven automated systems can deliver highly personalized experiences addressing a variety of customer service requests including opening new accounts, answering questions about existing accounts, providing assistance with investment and trades, reporting lost or stolen cards and helping with fraud detection.

The future of customer experience starts today

Chatbots are quickly evolving into digital avatars that can provide an omnichannel experience for your customers, delivering personalized and consistent service regardless of a customer's preferred channel. Digital avatars can be deployed on ATM machines to process transactions for customers. They may be used in kiosks at bank branches to answer questions and alleviate the burden on bank tellers. Avatars can exist in the [metaverse](#) to support your bank's presence in virtual worlds. They can also be used in mobile applications and on the web to help customers with transactions, questions and recommendations.

Digital avatars require a plethora of AI models to be effective, all of which can be powered by the Dell + NVIDIA AI Platform for Banking.



Activate AI-empowered banking outcomes for all your use cases

Your customers expect exceptional, seamless omnichannel experiences and innovative products at a good value. You need to generate more opportunities from your data and ensure secure and compliant operations while reducing total cost of ownership (TCO). AI, HPC and analytics technologies can automate workflows and quickly scrutinize large volumes of data to help you meet these goals. Dell Technologies solutions for AI, built in partnership with NVIDIA, provide the flexibility, ease of use, and cost profile you need to be successful with training and hosting AI models for identity verification, fraud detection and digital assistants. By taking advantage of the Dell Technologies knowledge, expertise and partnerships, you can achieve digital transformation that delivers on AI-empowered banking outcomes.

⁷ Forrester, [The Business Impact Of Investing In Experience](#), June 2021.

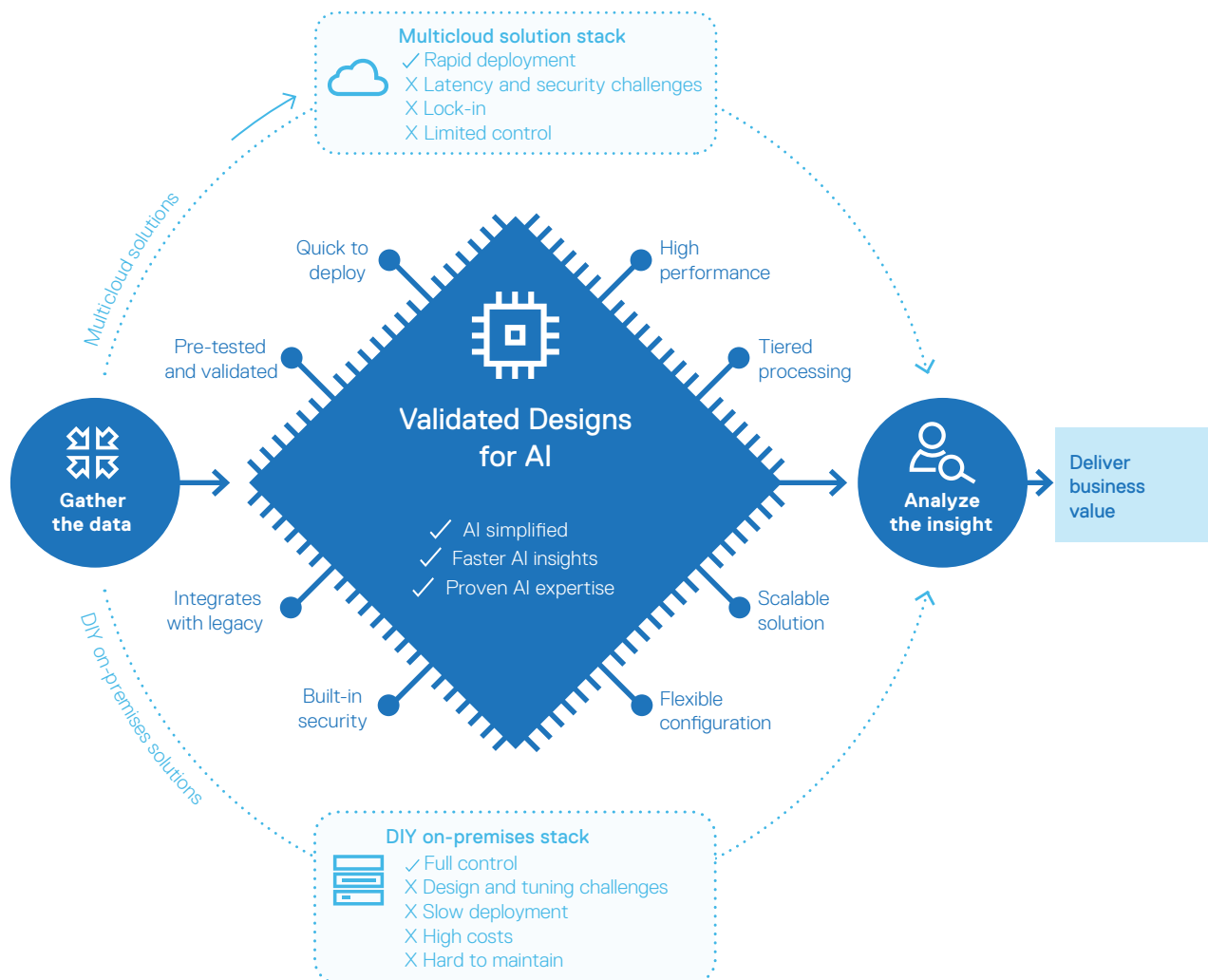
⁸ Hubtype, [Banking Chatbots 101: Benefits, Use Cases, and Statistics](#), May 2022.

Simplify and speed AI with Dell Validated Designs

To activate AI-empowered banking outcomes, you need a reliable infrastructure including compute, storage and networking optimized to work together to keep AI teams innovating, not waiting. But designing and deploying an AI system with the performance and scalability required can be complex. The key to successful AI implementation is adopting AI as a platform: a full-stack solution that includes both hardware and software.

Dell Validated Designs for AI accelerate time to value with solutions designed for the intelligent bank using Dell and NVIDIA proven AI expertise. Joint engineering and validation make it quick and easy to deploy an AI solution optimized to accelerate AI initiatives for faster AI insights.

Accelerate time to value with solutions designed for the intelligent bank



Know what's under the hood

Designing and deploying a system for AI, HPC and analytics with the performance and scalability required can be complex. Dell Technologies has invested to create a portfolio of Validated Designs to speed time to results with the confidence of engineering validation, enabling business without boundaries. These proven solutions have been optimized, tuned and tested for a variety of key use cases. They include the servers, storage, networking, software and services that have been proven in our labs and in customer deployments. Plus, the modular building block approach provides a customizable yet validated method for deploying new clusters or upgrading existing systems. These scalable systems are built with best-in-class solution stacks to deliver faster time to production, better performance and easier ability to scale. This allows your IT team to gather the data, process and analyze the data to garner insights and make decisions that deliver business value.

- **Validated Designs for AI** help make AI simpler with designs enabling you to get faster, deeper insights delivered with proven AI expertise.
- **Validated Designs for HPC** are scalable systems tested and tuned for specific vertical market applications such as financial services, life sciences, digital manufacturing and research.
- **Validated Designs for Analytics** drive competitive advantage with solutions designed to simplify deployment of analytics projects and protect your data.

You can save time and effort with architected, tested and validated solutions bringing together powerful and scalable compute, networking and storage designed to dynamically fit your needs based on specific use cases such as MLOps, Conversational AI and AutoML. For more information on specific Validated Designs, please refer to the [appendix](#).

NVIDIA AI Enterprise is an end-to-end, cloud-native suite of AI and data analytics software. It is optimized, certified and supported by NVIDIA to run on Dell PowerEdge NVIDIA-Certified Systems™. NVIDIA AI Enterprise includes key enabling technologies and software from NVIDIA for rapid deployment, management and scaling of AI workloads in the modern hybrid cloud. NVIDIA licenses and supports NVIDIA AI Enterprise. It includes:

- AI and data science frameworks such as Tensorflow™, PyTorch®, RAPIDS, NVIDIA TensorRT and NVIDIA Triton™ Inference Server.
- Cloud-native deployment components such as NVIDIA GPU operators for deploying and scaling GPU resources in Kubernetes®.
- Infrastructure optimization software such as drivers and libraries that optimize GPU resources.

Most AI workloads are modern applications that run on a container orchestration platform like Kubernetes. Validated Designs provide options for running AI workloads with Kubernetes. For example:

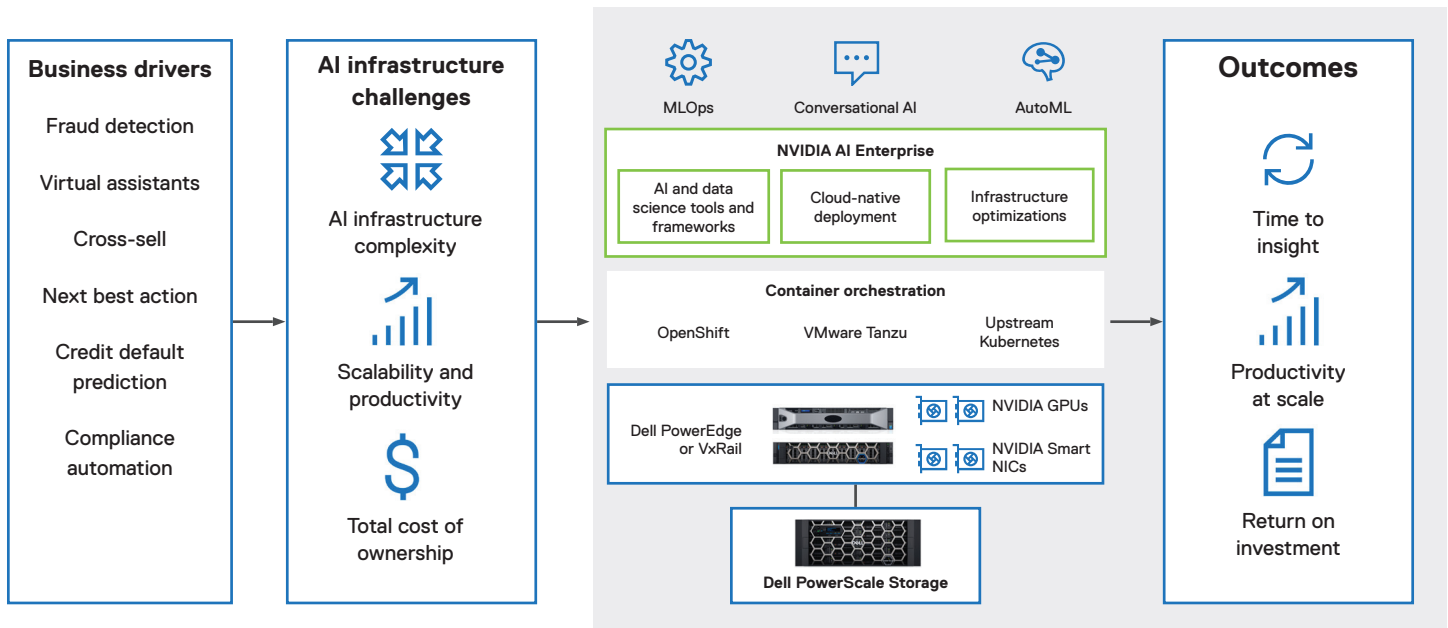
- VMware® vSphere® with Tanzu™ enables administrators to transform vSphere into a platform for running Kubernetes workloads natively on the hypervisor layer. When enabled on



a vSphere cluster, vSphere with Tanzu provides the capability to run Kubernetes workloads directly on ESXi™ hosts and to create upstream Kubernetes clusters in dedicated resource pools. vSphere administrators can enable existing vSphere clusters for workload management, to create a Tanzu Kubernetes cluster in the ESXi hosts that are part of the cluster. The Tanzu Kubernetes cluster is a full distribution of the open-source Kubernetes container orchestration platform that is built, signed and supported by VMware. Tanzu Kubernetes Grid (TKG) Service provisions and operates Tanzu Kubernetes cluster on vSphere NVIDIA GPUs through NVIDIA AI Enterprise. Virtual GPUs are automatically provisioned and configured on the Tanzu Kubernetes cluster worker nodes and made available to AI workload containers.

- Red Hat® OpenShift® automates application workload operations based on the upstream Kubernetes project. OpenShift uses the OpenShift Kubernetes Engine, an enterprise-grade Kubernetes distribution, to provide production-oriented container and workload automation.

Solution configuration overview



Breaking down the ROI

Having tailor-made systems with the optimized compute power necessary to operate AI workloads decreases the amount of time needed to solve problems, work on projects, train models and provide service.

Organizations using Dell Validated Designs for AI experience 20% faster time to value and benefits of \$55.76 million over three years versus costs of \$19.73 million, adding up to a net present value (NPV) of \$36.03 million and an ROI of 183%. Validated Designs for AI are more efficient than previous solutions, which reduces operational costs for AI, resulting in \$4.7 million in savings over three years. This includes reducing the required floor space and energy spend by 30% and 35%, respectively.⁹

But these numbers don't tell the whole story. Less fraud means more data remains secure, customers' livelihoods are protected, employee productivity remains high and the organization's financial health is protected.

Unquantified benefits of Validated Designs for AI include:

- Increased employee satisfaction
- Greater success in recruiting and retaining data scientists
- Improved customer reputation
- Environmental impacts

"The hardware would usually be the problem, but our partnership with Dell Technologies has taken that off the table. With Validated Designs for AI, instead of focusing on the setup, we can focus on developing our AI solution and delivering business value."

—CTO, Software⁹

⁹ Forrester, [The Total Economic Impact™ Of Dell Validated Designs For AI](#), August 2022.

Establish a flexible foundation with Dell APEX-as-a-Service solutions

Modernize for AI by expanding your use of cloud

Cloud computing is a powerful force with the capability to truly transform operations. However, many banks have yet to fully realize its potential. Most use cloud to make tactical changes to existing infrastructure, processes and applications or to add capabilities to on-premises environments without fundamentally altering operational models. Modernizing to expand the use of cloud is a critical step for financial services organizations seeking to advance their AI capabilities.

More than two-thirds of banking and investment services firms expect to increase their cloud spending in the next 12 months,¹⁰ working with a single vendor to reduce compliance risks. Their primary goals are to improve efficiency, increase data security and modernize IT.

¹⁰ Gartner, [Application modernization is the most common identified priority for end-user cloud adoption in banking and investment services](#), August 2021.

¹¹ Gartner, 2022 Outlook Perspective: Enterprise IT Spending Forecast for the Banking and Investment Services Market, June 2022.

Key cloud trends

› **AI workloads are moving to the cloud.** Though AI is mostly run on-premises today, many organizations are shifting toward an environment that utilizes both cloud and on-premises options to improve infrastructure utilization, scalability and performance.

› **High costs and other challenges remain.** Cloud migration is complex and requires careful consideration of security and compliance issues, team skills, application requirements and the costs and benefits of moving each workload to the cloud.

› **Hybrid cloud is the new reality.** Despite the advantages of cloud, cost, performance and security concerns will keep certain workloads on-premises, resulting in hybrid cloud operating environments that allow teams to optimize AI workload placement based on a variety of factors.

Cloud adoption spend focus

38% **34%** **32%**

Improved
efficiency¹¹



Increased data
security¹¹



IT
modernization¹¹



Dell APEX

At Dell Technologies, our goal is to support your digital transformation as you expand your use of AI. Infrastructure-as-a-Service (IaaS) offerings from [Dell APEX](#) ease or eliminate infrastructure management tasks to enable business and IT outcomes today and help you prepare for emerging application scenarios as they arise. Dell APEX makes it easier for you to get more value from Dell Technologies innovation by taking advantage of an as-a-Service cloud model, where and when you need it, whether for infrastructure, platforms or next-level solutions.

Dell APEX agility lets you react quickly to capture new opportunities and keep technology aligned with business needs. And with Dell APEX, you're in full control, empowering you to minimize risk and maximize performance, all on your terms. What does this mean? Get the benefits of as-a-Service for IT:

- Infrastructure is hosted in your data center, colocation or edge location.
- Systems are operated by you, managed for you.
- You can scale as needed.
- Users get fast access to self-service resources.
- The bank benefits from pay-per-use flexibility.

[Dell APEX Cloud Services](#) are built to meet the challenges of the multicloud era, giving you a more secure and consistent experience with the industry-leading capabilities and performance that mission-critical applications and databases require.

Dell APEX

Storage	Cyber and data protection	Compute and HCI	Solutions	Custom	More
Dell APEX Data Storage Services	Dell APEX Cyber Recovery Services	Dell APEX Hybrid Cloud	Dell APEX Containers for Red Hat OpenShift	Dell APEX Flex on Demand	
	Dell APEX Backup Services	Dell APEX Private Cloud	Dell APEX High Performance Computing	Dell APEX Data Center Utility	

We deliver a simple cloud experience that fits your bank's specific needs – and that includes data protection needs with trusted cybersecurity. Looking at the combined portfolio, you can see how our cloud services satisfy a range of data requirements. And, as opposed to being an afterthought, cyber-recovery is at the forefront of our service offerings so you can adopt a modern cyber-resilient approach from a partner focused on integrity.

In addition, **Dell APEX Custom Solutions** enable IT to create its own on-demand environment with preferred infrastructure. [Dell APEX Flex on Demand](#) allows you to pay for storage, data protection, hyperconverged infrastructure (HCI), servers and Dell Professional Services as you use them, with immediate access to buffer capacity. It provides you with a flexible option for acquiring storage and compute capacity without a large capital outlay up front, giving you the ability to avoid overprovisioning storage and compute functionality and to easily scale up to manage unpredictable growth, workload bursts and temporary changes in IT infrastructure requirements.

[Dell APEX Data Center Utility](#) aligns costs directly to usage, allowing you to maximize your scaling flexibility while only paying for what you use. It provides the Dell market-leading product portfolio coupled with professional services and support to fully manage your data center and its operations in a simple, single-invoice monthly payment based on your actual usage.

From compute to data storage and data protection with cyber resiliency to custom solutions, all are available today in various ways and locations to fit your needs, including on-premises, cloud-adjacent or cloud-resident environments with supplemental professional services that provide the next-level integration, migration, residency and more.

Extending the value of your Dell APEX solution with Professional Services

Dell portfolio of value-added professional services for Dell APEX helps you optimize your operational model so you and your team can focus on innovation. Simplify your operations by:

- Keeping production workloads accessible with Dell migration.
- Speeding IT delivery requirements with automated provisioning.
- Simplifying operations by integrating workflows into an IT service portal.
- Employing residency services to act alongside your team to extend your skill sets and get the most out of your Dell APEX operations.

Partner with Dell Technologies and NVIDIA experts

HPC & AI Innovation Lab

The [HPC & AI Innovation Lab](#) in Austin, Texas, is our flagship innovation center. Housed in a 13,000 square foot data center, it gives you access to thousands of Dell servers, three powerful HPC clusters and sophisticated storage and network systems. It's staffed by a dedicated group of computer scientists, engineers and subject matter experts who actively partner and collaborate with customers and other members of the community. The team engineers AI solutions, tests new and emerging technologies and shares expertise, including performance results and best practices.

HPC & AI Centers of Excellence

Dell worldwide [HPC & AI Centers of Excellence](#) provide thought leadership, test new technologies and share best practices. They maintain local industry partnerships and have direct access to Dell and other technology creators to incorporate your feedback and needs into their roadmaps. Through collaboration, these Centers of Excellence provide a network of resources based on the wide ranging know how and experience in the community.

61%

of organizations are interested in shifting IT to consumption-based/as-a-Service models¹²

39%

lower annual infrastructure costs with Dell APEX¹²

64%

reduction in time to deploy new capacity with Dell APEX¹²





Services and financing for AI

Along your AI journey, Dell Technologies is with you every step of the way, linking people, processes and technology to accelerate innovation and enable optimal business outcomes.

[Consulting Services](#) help you create a competitive advantage for your business. Our expert consultants work with companies at all stages to help you plan, implement and optimize solutions that enable you to unlock your data capital and support AI.

[Deployment Services](#) help you streamline complexity and bring new IT investments online as quickly as possible. Leverage our 30 plus years of experience for efficient and reliable solution deployment to accelerate adoption and return on investment while freeing IT staff for more strategic work.

[Support Services](#) driven by AI will change the way you think about support with smart, groundbreaking technology backed by experts to help you maximize productivity, uptime and convenience. Experience more than fast problem resolution — our AI engine proactively detects and prevents issues before they impact performance.

[Payment Solutions](#) from Dell Financial Services help you maximize your IT budget and get the technology you need today. Our portfolio includes traditional leasing and financing options, as well as advanced flexible consumption products.

Dell [APEX](#) offers simple and consistent cloud experiences, delivered as-a-Service. Provision quickly, scale on-demand and pay as you go for Dell Technologies innovation, unlocking the flexibility you need to adapt and thrive.

[Managed Services](#) can help reduce the cost, complexity and risk of managing IT so you can focus your resources on digital innovation and transformation while our experts help optimize your IT operations and investment.

[Residency Services](#) provide the expertise needed to drive effective IT transformation and keep IT infrastructure running at its peak. Resident experts work tirelessly to address challenges and requirements, with the ability to adjust as priorities shift.

Activate better banking outcomes with AI

Banks are at a crossroads. You need to achieve speed, agility and flexibility to achieve your goals. Yet you must continue to manage scale, compliance and regulatory requirements while exceeding customer needs.

Working together, Dell Technologies and NVIDIA help you prepare your organization for solution evolution by developing foundational data science and AI/ML capabilities to support new types and forms of data, models and partnership, realizing the vision of the AI-empowered bank.

[Learn More](#)
Dell.com/FinancialServicesIndustry





Appendix

Data center-ready enterprise MLOps with DOMINO

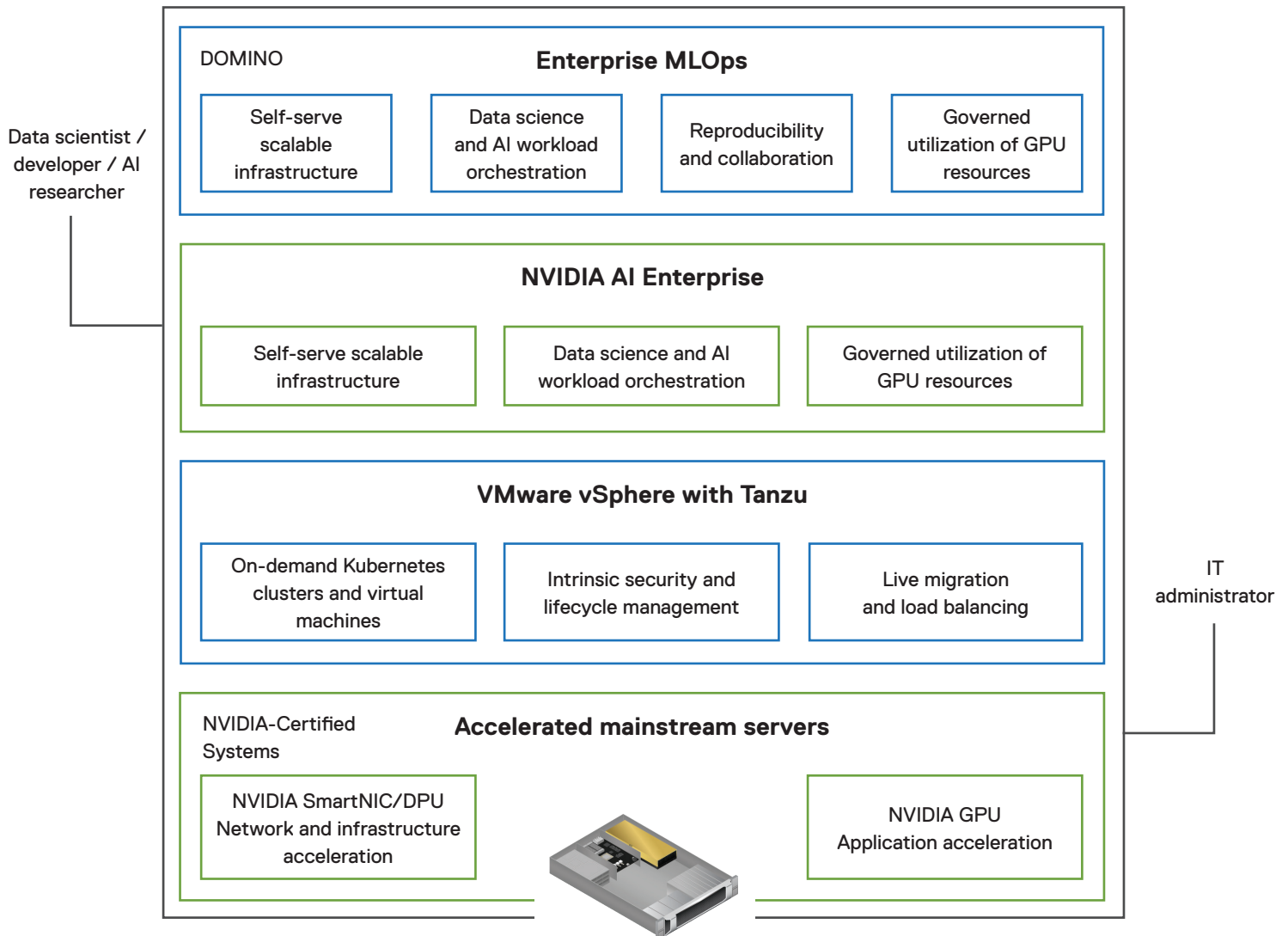
AI fuels digital transformation for many companies, allowing them to automate tasks, streamline decision processes and deploy use cases that were unachievable 10 years ago. However, assembling the right mix of people, data sources, computing resources and team structure to develop intelligent software applications with good return on investment is a new skill set for many organizations.

One of the most complex tasks for getting AI workloads into production is model development. Data scientists need to be able to streamline development to deployment by using a platform that accelerates the data science lifecycle with open access to tools, compute and data. Enterprises require an open and flexible data science platform that automates tasks while allowing teams to collaborate efficiently, enabling the use cases of today while not locking out those of tomorrow.

The DOMINO® Data Lab Enterprise MLOps platform is designed to address the unique needs of large enterprises who want to get more models into production faster. The solution supports open, collaborative, reproducible model development, training and management free of DevOps constraints — powered by efficient, end-to-end compute. DOMINO's validation for NVIDIA AI Enterprise pairs the Enterprise MLOps benefits of workload orchestration, self-serve infrastructure and collaboration with cost-effective scale from virtualization on accelerated Dell PowerEdge servers.

Dell Technologies is pushing the boundaries of performance for AI workloads with scalable Validated Designs for AI. These standardized building blocks simplify design and speed configuration and ordering of solutions that have been rigorously tested and tuned for AI applications. Dell Technologies and DOMINO have partnered to validate a solution that provides an optimal environment to run DOMINO Data Science Platform, enhancing performance for AI and ML workloads that are critical for advancing business objectives.

For added flexibility, the Dell Validated Design for DOMINO Data Science Platform uses a flexible approach to system design, where individual building blocks can be combined to build a system optimized specifically for unique workloads and use cases. By providing architecture guidance that accelerates deployment, saves time and conserves resources, Dell Technologies and DOMINO Data Lab can help you set up an optimized environment in weeks rather than months.



Date center-ready enterprise MLOps with cnvrg.io

Organizations today use AI to support business intelligence goals and solve problems. ML is the most critical enabler of AI, employing algorithms and learning models to parse large data sets. Image and speech recognition are two examples of applications where employing ML can improve AI far faster than human analysis can.

But data scientists working with ML spend an inordinate amount of time designing, configuring and testing ML platforms, causing these highly trained specialists to spend too little time working with the data and models, which is their primary specialty. By using MLOps — a term inspired by the popular DevOps model for application development — organizations investing in ML can more easily automate the continuous training and deployment of ML models at scale. MLOps adapts DevOps principles and practices to the ML workflow.

Combining development and operations in DevOps creates a more standardized and accelerated methodology for application development and deployment. Similarly, by automating the complexity and variability of the ML process, MLOps helps lead to far more reproducible, testable and evolvable ML environments.

Standardized, predictable and manageable ML deployments can drive the launch of new capabilities, discoveries, and services, enabling an organization to derive more insights and value from the data it collects. These insights and values are the goal of MLOps.

Available either as self-hosted software or as the Metacloud SaaS offering, cnvrg.io delivers a full-stack MLOps platform that helps simplify continuous training and deployment of AI and ML models. With cnvrg.io, organizations can automate end-to-end ML pipelines at scale in all environments, whether on-premises or across clouds. cnvrg.io makes it easy to place training or inferencing workloads on CPUs, GPUs, TPUs and other specialized accelerators depending on the cost and performance requirements.



Regardless of the components in the pipeline, the result is a single end-to-end flow designed to maximize workload performance, optimizing with the right hardware and processing elements beneath each stage in the flow. ML jobs can be launched on-demand, regardless of the underlying compute and storage elements.

Whether from the command line, a software development kit (SDK) or an intuitive visual interface, cnvrg.io provides access to all models, code and data sets that can be run across an organization's compute and storage resources. Utilization and efficiency are boosted as data scientists can aggregate and use the best components, optimized for the job, and then orchestrate those flows, all from a unified graphical interface. Through native Kubernetes pod and cluster orchestration and cnvrg.io internal job scheduler, both cloud and on-premises resources can be easily scaled to meet the computational and storage needs of an organization's ML workloads.

All these capabilities help remove friction and latency from the data science process, getting models out of the lab and into production quicker and reducing time to value for of the data. By removing much of the underlying complexity from the model, data scientists can spend more time delivering insight and spend less time dealing with configuration and testing. With cnvrg.io, enterprises can apply MLOps for continuous training and deployment of ML in the way that DevOps principles enable CI/CD for traditional IT workloads. Dell Technologies has worked closely with cnvrg.io to deliver MLOps for AI/ML adopters through a jointly engineered and tested solution to help organizations capitalize on the benefits of MLOps for ML and AI workloads. Developers get a cloud-like self-service experience for choosing compute and storage resources from market leaders like Dell Technologies.

More accessible ML using AutoML with H2O.ai

Organizations across a wide range of industries are taking note of the ability of AI to help them reduce risk, create value and uncover new opportunities. And AI is driven by data — how it's analyzed, managed, stored, processed, protected and leveraged. Emerging technologies, such as Internet of Things (IoT) and mobile technologies and applications, are creating data at speeds and volumes never imagined. This data needs AI to be converted into the fuel that enables organizations to make better decisions, faster.

However, deploying systems capable of running AI workloads can be complex, requiring extensive integration and testing of the hardware and software. That's why Dell Technologies is working with H2O.ai to bring you engineering-validated designs for H2O.ai.

H2O® is a fully open-source distributed in-memory AI and ML platform with linear scalability. Hundreds of thousands of data scientists in more than 18,000 global enterprises trust H2O because it supports some of the most widely used statistical and ML algorithms — including gradient boosted machines, generalized linear models, DL and more. H2O is also incredibly flexible. It works on bare metal, with existing Apache® Hadoop® or Apache Spark® clusters. It can ingest data directly from HDFS, Spark, S3, Microsoft® Azure® Data Lake or other data sources into its in-memory distributed key value store. In addition, H2O has industry-leading AutoML functionality that automatically runs through algorithms and their hyperparameters to produce a leaderboard of the best performing models.

H2O takes advantage of the computing power of distributed systems and in-memory computing to accelerate ML using its industry parallelized algorithms, which take advantage of fine-grained in-memory MapReduce. Data scientists can quickly and easily deploy models into production with Java® (POJO) and binary formats (MOJO). Dell Technologies partners with H2O.ai to create Validated Designs specifically for H2O software, to enhance performance for AI and ML workloads that are critical for advancing business objectives. For added flexibility, Validated Designs for H2O.ai use a flexible building block approach to system design, where individual building blocks can be combined to build a system that's optimized specifically for your unique workloads and use cases.

Supplemental information

Benefits of using NVIDIA GPUs in Dell solutions

Deploying AI capabilities requires a scalable, resilient and adaptable set of core technology components. NVIDIA GPUs enable financial firms to calculate pricing, risk for complex options/over-the-counter (OTC) derivatives, and other high frequency trading (HFT) and ultra-high frequency trading (UHFT) algorithms in seconds instead of minutes or hours by using parallel processing. NVIDIA GPUs, which include the V100, A100 and T4, are designed to drive immense productivity in quantitative research at AI-empowered financial services firms. Libraries such as NVIDIA CUDA® are designed for HPC-AI calculators.

NVIDIA NVLink™ is a high-speed, direct GPU-to-GPU interconnect essential for artificial neural nets training and inferencing. NVSwitch takes interconnectivity to the next level by incorporating multiple NVLinks to provide all-to-all GPU communication within a single node, useful for neural net algorithm tasks where multiple GPUs can be clustered to perform parallel computations.

HPC-AI scaling approaches that make use of clustered compute using ultralow latency connectivity are seeing increased adoption. To derive optimal benefits, IT organizations must make use of a combination of NVIDIA NVLink, 10GbE and NVIDIA InfiniBand® for various aspects of the cluster such as north-south data access, east-west node-to-node traffic and north-south external access.

NVIDIA CUDA is an SDK and application programming interface (API) that allows the use of the C++ programming language to code algorithms for execution on GPUs. The parallel computing platform and programming model dramatically increase the performance of applications using GPU-accelerated libraries and make using a GPU for general-purpose computing simple.

CUDA Deep Neural Network (cuDNN) library, a GPU-accelerated library of primitives for deep neural networks, includes components such as cuBLAS, and cuSPARSE for standard basic linear algebra subroutines and cuSOLVER to accelerate linear optimization applications. In addition, NCCL Parallel Algorithm Library for Quantitative Research implements multi-GPU and multinode collective communication primitives. Thrust provides a flexible developer interface for parallel algorithms and data structures along with CUDA-X — NVIDIA CUDA-X. Developing these applications requires a robust programming environment with highly optimized, domain-specific libraries. Further, the RAPIDS suite of software libraries built on CUDA-X AI shows parallelism and high-bandwidth memory speed through user-friendly Python® interfaces. With NVIDIA RAPIDS, it is possible to write HPC applications in NumPy instead of C++ as Python becomes the de facto language of choice for accelerated analytics.



Copyright © 2023 Dell Inc. or its subsidiaries. All Rights Reserved. Dell Technologies, Dell and other trademarks are trademarks of Dell Inc. or its subsidiaries. NVIDIA®, NVIDIA, NVIDIA-Certified Systems™, NVIDIA Triton™, NVIDIA CUDA®, NVIDIA NVLink™, and NVIDIA InfiniBand® are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. Mastercard® is a registered trademark or service mark of Mastercard or its subsidiaries in the United States. TensorFlow™ and any related marks are trademarks of Google Inc. PyTorch® is a trademark or registered trademark of PyTorch or PyTorch's licensors. Kubernetes® is a registered trademark of The Linux Foundation. VMware® is a registered trademark or trademark of VMware, Inc. in the United States and other jurisdictions. Red Hat® and Red Hat OpenShift® are registered trademarks of Red Hat, Inc. in the United States and other countries. DOMINO® is a registered trademark of DOMINO Data Lab, Inc. H2O® is a trademark of H2O.ai. Apache®, Apache Hadoop®, and Apache Spark® are trademarks of the Apache Software Foundation or its subsidiaries in Canada, the United States, and/or other countries. Microsoft® and Microsoft Azure® are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries. Java® is a registered trademark of Oracle and/or its affiliates. Python® is a registered trademark of the Python Software Foundation. Other trademarks may be the property of their respective owners. Published in the USA 01/23 Whitepaper ai-banking-wp-101.

Dell Technologies believes the information in this document is accurate as of its publication date. The information is subject to change without notice.